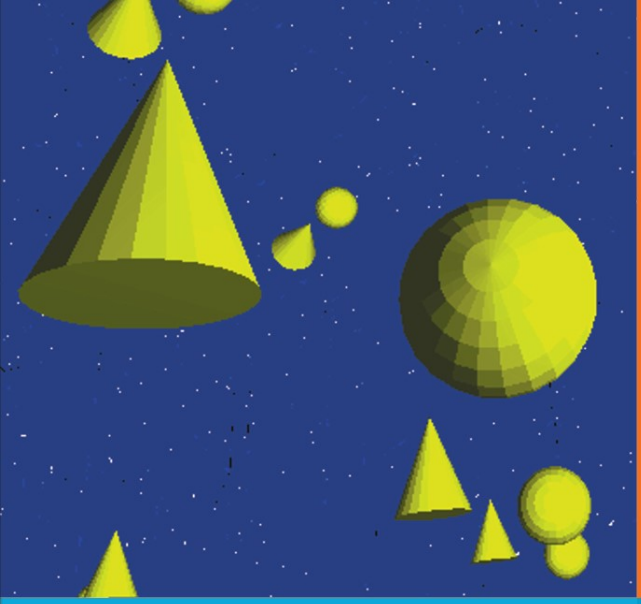


Max K. Agoston

# Computer Graphics and Geometric Modeling

## Mathematics

Springer

# Computer Graphics and Geometric Modeling

Max K. Agoston

# Computer Graphics and Geometric Modeling

**Mathematics**

Springer

Max K. Agoston, MA, MS, PhD
Cupertino, CA 95014, USA

British Library Cataloguing in Publication Data
Agoston, Max K.
Computer graphics and geometric modeling mathematics
1. Computer graphics 2. Geometry – Data processing 3. Computer-aided design
4. Computer graphics – Mathematics
I. Title
006.6
ISBN 1852338172

Library of Congress Cataloging-in-Publication Data

Agoston, Max K.
Computer graphics & geometric modeling / Max K. Agoston.
p. cm.
Includes bibliographical references and index.
Contents: Mathematics.
ISBN 1-85233-817-2 (alk. paper)
1. Computer graphics. 2. Geometry – Data processing. 3. Mathematical models.
4. CAD/CAM systems. I. Title.

T385.A395 2004
006.6 – dc22 2004049155



ISBN 1-85233-817-2
Springer is part of Springer Science+Business Media
springeronline.com


Printed in the United States of America



Typesetting: SNP Best-set Typesetter Ltd., Hong Kong
34/3830-543210 Printed on acid-free paper SPIN 10971444

# Preface

This book and [AgoM05] grew out of notes used to teach various types of computer graphics courses over a period of about 20 years. Having retired after a lifetime of teaching and research in mathematics and computer science, I finally had the time to finish these books. The goal of these books was to present a comprehensive overview of computer graphics as seen in the context of geometric modeling and the mathematics that is required to understand the material. The reason for two books is that there was too much material for one. The practical stuff and a description of the various algorithms and implementation issues that one runs into when writing a geometric modeling program ended up in [AgoM05], and the mathematical background for the underlying theory ended up here. I have always felt that understanding the mathematics behind computer graphics was just as important as the standard algorithms and implementation details that one gets to in such courses and included a fair amount of mathematics in my computer graphics courses.

Given the genesis of this book, the primary intended audience is readers who are interested in computer graphics or geometric modeling. The large amount of mathematics that is covered is explained by the fact that I wanted to provide a **complete** reference for all the mathematics relevant to geometric modeling. Although computer scientists may find sections of the book very abstract, everything that was included satisfied at least one of two criteria:

(1) It was important for some aspect of a geometric modeling program, or
(2) It provided helpful background material for something that might be used in such a program.

On the other hand, because the book contains only mathematics and is so broad in its coverage (it covers the basic definitions and main results from many fields in mathematics), it can also serve as a reference book for mathematics in general. It could in fact be used as an introduction to various topics in mathematics, such as topology (general, combinatorial, algebraic, and differential) and algebraic geometry.

Two goals were very important to me while writing this book. One was to thoroughly explain the mathematics and avoid a cookbook approach. The other was to make the material as self-contained as possible and to define and explain pretty much every technical term or concept that is used. With regard to the first goal, I have tried

very hard to present the mathematics in such a way that the reader will see the motivation for it and **understand** it. The book is aimed at those individuals who seek such understanding. Just learning a few formulas is not good enough. I have always appreciated books that tried to provide motivation for the material they were covering and have been especially frustrated by computer graphics books that throw the reader some formulas without explaining them. Furthermore, the more mathematics that one knows, the less likely it is that one will end up reinventing something. The success or failure of this book should be judged on how much understanding of the mathematics the reader got, along with whether or not the major topics were covered adequately.

To accomplish the goal of motivating **all** of the mathematics needed for geometric modeling in one book, even if it is large, is not easy and is impossible to do from scratch. At some places in this book, because of space constraints, few details are provided and I can only give references. Note that I always have the nonexpert in mind. The idea is that those readers who are not experts in a particular field should at least be shown a road map for that field. This road map should organize the material in a logical manner that is as easy to understand and as motivated as possible. It should lay out the important results and indicate what one would have to learn if one wanted to study the field in more detail. For a really in-depth study of most of the major topics that we cover, the reader will have to consult the references.

Another of my goals was to state everything **absolutely** correctly and not to make statements that are only approximately correct. This is one reason why the book is so long. Occasionally, I had to digress a little or add material to the appendices in order to define some concepts or state some theorems because, even though they did not play a major role, they were nevertheless referred to either here or in [AgoM05]. In those cases involving more advanced material where there is no space to really get into the subject, I at least try to explain it as simply and intuitively as possible. One example of this is with respect to the Lebesque integral that is referred to in Chapter 21 of [AgoM05], which forced the inclusion of Section D.4. Actually, the Lebesgue integral is also the only example of where a concept was not defined.

Not all theorems stated in this book are proved, but at least I try to point out any potential problems to the reader and give references to where the details can be found in those cases where proofs are omitted, if so desired. Proofs themselves are not given for their own sake. Rather, they should be thought of more as examples because they typically add insight to the subject matter. Although someone making a superficial pass over the mathematical topics covered in the book might get the impression that there is mathematics that has little relevance to geometric modeling, that is not the case. **Every** bit of mathematics in this book and its appendices is used or referred to somewhere here or in [AgoM05]. Sometimes defining a concept involved having to define something else first and so on. I was not trying to teach mathematics for its own interesting sake, but only in so far as it is relevant to geometric modeling, or at least potentially relevant. When I say "potentially," I am thinking of such topics as algebraic and differential topology that currently appear in only minimal ways in modeling systems but obviously will some day play a more central role.

It is assumed that the reader has had minimally three semesters of calculus and a course on linear algebra. An additional course on advanced calculus and modern algebra would be ideal. The role of Appendices B–F is to summarize what is assumed. They consist mainly of definitions and statements of results with essentially no expla-

nations. The reason for including them is, as stated earlier, to be self-contained. Readers may have learned the material at some point but forgotten the details, such as a definition or the precise statement of a theorem. A reader who does not understand some particular material in the appendices may not understand the discussion at those places in the book where it is used. The biggest of the appendices is Appendix B, which consists of material from modern algebra. This appendix is needed for Chapters 7, 8, and 10, although not that much of it is needed for Chapters 7 and 8. Only Chapter 10 on algebraic geometry needs a lot of that background. This is the one place where using this text in the context of a course would be a big advantage over reading the material on one's own because an instructor who knows the material would actually be able to explain the important parts of it quite easily and quickly even to students who have **not** had a prior course on modern algebra. The actual applications of Chapter 10 to geometric modeling do not require that much knowledge if one skips over the background and proofs of the theorems that lead up to them. Hopefully, however, the reader with a minimal mathematics background will be reduced to simply learning "formulas" in only a few places in this book.

The extensive material on topology, in particular algebraic and differential topology, has heretofore not been found in books directed toward geometric modeling. Although this subject is slowly entering the field, its coming has been slow. Probably the two main reasons for this are that computers are only now getting to be powerful enough to be able to handle the complicated computations, and the material involves exceptionally advanced mathematics that even mathematics majors would normally not see until graduate school. It is not very surprising therefore that in cases like this most of the advancement here will probably come from mathematicians who either switch their research interest to computer science or who want to use computers to make advances in their field. Having said that though, I also strongly feel that there is much that can be explained to a nontopologist, and Chapters 6–8 are an attempt to do this. A similar comment applies to the algebraic geometry in Chapter 10. It is because of my emphasis on explaining things that I suggested earlier that mathematics students could also use this book to learn about this material, not just computer scientists.

With regard to the bibliography, it is fairly small because the book is not addressed to mathematicians per se. This meant that many good but **advanced** references that I could have given, but whose intended audience is research mathematicians, are omitted. This lack of completeness is partially compensated by the fact that additional references can be found in the references that are given.

The numbering of items in this book uses the following format: x.y.z refers to item number z in section y of chapter x. For example, Theorem 6.5.7 refers to the seventh item of type theorem, proposition, lemma, or example in section 5 of Chapter 6. Algorithm 10.11.1 refers to the first algorithm in Section 11 of Chapter 10. Tables are numbered like algorithms. Figures are numbered by chapter, so that Figure 9.21 refers to the twenty-first figure in Chapter 9. Exercises at the end of chapters are numbered by section.

Cupertino, California Max K. Agoston

# Contents

CHAPTER 1

# Linear Algebra Topics

## 1.1 Introduction

This chapter assumes a basic knowledge and familiarity of linear algebra that is roughly equivalent to what one would get from an introductory course in the subject. See Appendix B and C for the necessary background material. In particular, we assume the reader is familiar with the vector space structure of n-dimensional Euclidean space $\mathbf{R}^n$ and its dot product and associated distance function. The object of this chapter is to discuss some important topics that may not have been emphasized or even covered in an introductory linear algebra course. Those readers with a weak background in abstract linear algebra and who have dealt with vectors mostly in the context of analytic geometry or calculus topics in $\mathbf{R}^2$ or $\mathbf{R}^3$ will also get a flavor of the beauty of a coordinate-free approach to vectors. Proofs should not be skipped because they provide more practice and insight into the geometry of vectors. The fact is that a good understanding of (abstract) linear algebra and the ability to apply it is essential for solving many problems in computer graphics (and mathematics).

As in other places in this book we have tried to avoid generality for generality's sake. By and large, the reader can interpret everything in the context of subspaces of $\mathbf{R}^n$; however, there are parts in this chapter where it was worthwhile to phrase the discussion more generally. We sometimes talk about inner product spaces, rather than just sticking to $\mathbf{R}^n$ and its dot product, and talk about vector spaces over other fields, the complex numbers $\mathbf{C}$ in particular. This was done in order to emphasize the general nature of the aspect at hand, so that irrelevant details do not hide what is important. Vector spaces over the complex numbers will be important in later chapters.

Geometry is concerned with lots of different types of spaces. This chapter is about the simplest of these, namely, the linear ones, and some related topics. Hopefully, much of the material that is covered is review except that we shall approach the subject here, like in many other places, with a vector approach. Sections 1.2–1.5 review the definition and basic properties of k-dimensional planes in $\mathbf{R}^n$. We also look at the abstract definition of angle and some important concepts related to orthogonality, such as that of the orthogonal projection of a vector. Next, in Sections 1.6 and 1.7 we discuss the extremely important concepts of orientation and convexity.

Some basic results on the diagonalization of maps and matrices in Section 1.8 lead to a discussion of bilinear maps and quadratic forms in Section 1.9. Section 1.10 describes a general version of the three-dimensional cross product. Finally, Section 1.11 defines the generalized inverse of a transformation and matrix along with several applications.

## 1.2 Lines

Our first goal in this chapter is to characterize linear subspaces of Euclidean space and summarize some basic facts about them. There is not much to say about points, the 0-dimensional linear subspaces, but the one-dimensional subspaces, namely, "straight" lines, are a special case that is worth looking at separately.

First of all, let us consider lines in the plane. The usual definition of a line in this case is as the set of solutions to a linear equation.

**Definition.** (The **equation** definition of a line in the plane) Any set **L** in $\mathbf{R}^2$ of the form

$$\{(x,y) \mid ax + by = c, (a,b) \neq (0,0)\}, \tag{1.1}$$

where a, b, and c are fixed real constants, is called a *line*. If a = 0, then the line is called a *horizontal line*. If b = 0, then the line is called a *vertical line*. If b ≠ 0, then –a/b is called the *slope* of the line.

Although an equation defines a unique line, the equation itself is not uniquely defined by a line. One can multiply the equation for a line by any nonzero constant and the resultant equation will still define the same line. See Exercise 1.2.1.

The particular form of the equation in our definition for a line is a good one from a theoretical point of view, but for the sake of completeness we list several other well-known forms that are often more convenient.

The **slope-intercept form:** The line with slope m and y-intercept (0,b) is defined by

$$y = mx + b. \tag{1.2}$$

The **point-slope form:** The line through the point $(x_1,y_1)$ with slope m is defined by

$$y - y_1 = m(x - x_1). \tag{1.3}$$

The **two-point form:** The line through two distinct points $(x_1,y_1)$ and $(x_2,y_2)$ is defined by

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1). \tag{1.4}$$

Note that equations (1.2) and (1.3) above apply **only** to **nonvertical** lines.

When one wants to define lines in higher dimensions, then one can no longer use a single equation and so we now give an alternative definition that works in all dimen-

sions. It is based on the intuitive geometric idea that a line is defined by a point and a direction.

**Definition.** (The **point-direction-vector** definition of a line) Any subset **L** of $\mathbf{R}^n$ of the form

$$\{\mathbf{p} + t\mathbf{v} \mid t \in \mathbf{R}\}, \tag{1.5a}$$

where **p** is a fixed point and **v** is a fixed nonzero vector in $\mathbf{R}^n$, is called a *line* (*through* ***p***). The vector **v** is called a *direction vector* for the line **L**. By considering the components of a typical point $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ in **L** separately, one gets equations

$$\begin{aligned} x_1 &= p_1 + tv_1 \\ x_2 &= p_2 + tv_2 \\ &\cdot \quad \cdot \quad \cdot \\ x_n &= p_n + tv_n,\ t \in \mathbf{R}, \end{aligned} \tag{1.5b}$$

that are called the *parametric equations for the line*.

In the case of the plane, it is easy to show that the two definitions of a line agree (Exercise 1.2.2). The definition based on the equation in (1.1) is an **implicit** definition, meaning that the object was defined by an equation, whereas the definition using (1.5a) is an **explicit** definition, meaning that the object was defined in terms of a **parameterization**. We can think of t as a time parameter and that we are walking along the line, being at the point $\mathbf{p} + t\mathbf{v}$ at time t.

Note that the direction vector for a line is not unique. Any nonzero multiple of **v** above would define the same line. Direction vectors are the analog of the slope of a line in higher dimensions.

**1.2.1. Example.** To describe the line **L** containing the points $\mathbf{p} = (0,2,3)$ and $\mathbf{q} = (-2,1,-1)$.

**Solution.** The vector $\mathbf{pq} = (-2,-1,-4)$ is a direction vector for **L** and so parametric equations for **L** are

$$\begin{aligned} x &= \quad -2t \\ y &= 2 - t \\ z &= 3 - 4t \end{aligned}$$

**1.2.2. Example.** Suppose that the parametric equations for two lines $\mathbf{L}_1$ and $\mathbf{L}_2$ are:

$$\begin{aligned} \mathbf{L}_1:\quad x &= 1 - t & \qquad \mathbf{L}_2:\quad x &= 2 + t \\ y &= 2 + t & y &= 1 - 2t \\ z &= -1 + t & z &= -2 + t \end{aligned} \tag{1.6}$$

Do the lines intersect?

**Solution.** We must solve the equations

$$\begin{aligned} 1-t &= 2+s \\ 2+t &= 1-2s \\ -1+t &= -2+s \end{aligned}$$

for s and t. The first two equations imply that $t = -1$ and $s = 0$. Since these two values also satisfy the third equation, the lines $\mathbf{L}_1$ and $\mathbf{L}_2$ intersect at the point (2,1,–2).

**Note.** A common mistake when trying to solve a problem like that in Example 1.2.2 is to use the same variable for both s and t. Just because lines intersect does not mean that persons "walking" along the lines will get to the intersection point at the same "time."

**Definition.** Points are said to be *collinear* if they lie on the same line and *non-collinear*, otherwise.

**Definition.** Let $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$. The set

$$\{\mathbf{p} + t\mathbf{pq} \mid t \in [0,1]\} \tag{1.7}$$

is called the *segment from* $\boldsymbol{p}$ *to* $\boldsymbol{q}$ and is denoted by $[\mathbf{p},\mathbf{q}]$. The points of $[\mathbf{p},\mathbf{q}]$ are said to lie *between* $\mathbf{p}$ and $\mathbf{q}$.

Note that $[\mathbf{p},\mathbf{q}] = [\mathbf{q},\mathbf{p}]$ (Exercise 1.2.5). A segment basically generalizes the notion of a closed interval of the real line, which explains the notation, but the two concepts are **not** quite the same when $n = 1$ (Exercise 1.2.6). The following proposition gives a very useful alternative characterization of a segment.

**1.2.3. Proposition.** Let $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$. Then

$$[\mathbf{p},\mathbf{q}] = \left\{\mathbf{x} \in \mathbf{R}^n \mid |\mathbf{px}| + |\mathbf{xq}| = |\mathbf{pq}|\right\}. \tag{1.8}$$

**Proof.** Let

$$\mathbf{S} = \{\mathbf{x} \mid |\mathbf{px}| + |\mathbf{xq}| = |\mathbf{pq}|\}.$$

In order to show that $[\mathbf{p},\mathbf{q}] = \mathbf{S}$ we must prove the two inclusions $[\mathbf{p},\mathbf{q}] \subseteq \mathbf{S}$ and $\mathbf{S} \subseteq [\mathbf{p},\mathbf{q}]$.

To prove that $[\mathbf{p},\mathbf{q}] \subseteq \mathbf{S}$, let $\mathbf{x} \in [\mathbf{p},\mathbf{q}]$. Then $\mathbf{x} = \mathbf{p} + t\mathbf{pq}$ for some t with $0 \le t \le 1$. It follows that

$$|\mathbf{px}| + |\mathbf{xq}| = |t|\,|\mathbf{pq}| + |1-t|\,|\mathbf{pq}| = |\mathbf{pq}|,$$

so that $\mathbf{x} \in \mathbf{S}$.

To prove that $\mathbf{S} \subseteq [\mathbf{p},\mathbf{q}]$, let $\mathbf{x} \in \mathbf{S}$. Since $|\mathbf{px}| + |\mathbf{xq}| = |\mathbf{pq}| = |\mathbf{px} + \mathbf{xq}|$, the triangle inequality implies that the vectors $\mathbf{px}$ and $\mathbf{xq}$ are linearly dependent. Assume without loss of generality that $\mathbf{px} = t\mathbf{xq}$. Then

$$|t|\,|\mathbf{xq}| + |\mathbf{xq}| = |t\mathbf{xq} + \mathbf{xq}|.$$

In other words,

$$|t| + 1 = |t + 1| \tag{1.9}$$

It is easy to show that the only solutions to (1.9) are $0 \le t$. But the equation $\mathbf{px} = t\mathbf{xq}$ can be rewritten as

$$\mathbf{x} = \mathbf{p} + \frac{t}{1+t}\mathbf{pq}$$

which shows that $\mathbf{x} \in [\mathbf{p},\mathbf{q}]$ since $0 \le t/(1 + t) \le 1$.

The next proposition proves another fairly innocuous looking fact. It also plays a key role in the proofs of a number of future theorems.

**1.2.4. Proposition.** Let $\mathbf{p}$ be a point on a line $\mathbf{L}$. If $c > 0$, then there are two and only two points $\mathbf{x}$ on $\mathbf{L}$ that satisfy the equation $|\mathbf{px}| = c$.

**Proof.** Let $\mathbf{q}$ be a point on $\mathbf{L}$ distinct from $\mathbf{p}$. Then any point $\mathbf{x}$ on $\mathbf{L}$ has the form $\mathbf{x} = \mathbf{p} + s\mathbf{pq}$ and hence $c = |\mathbf{px}| = |s|\,|\mathbf{pq}|$. The only solutions to $|s| = c/|\mathbf{pq}|$ are $s = \pm t$, where $t = c/|\mathbf{pq}|$. In other words,

$$\mathbf{x} = \mathbf{p} + t\mathbf{pq} \quad \text{or} \quad \mathbf{x} = \mathbf{p} - t\mathbf{pq}$$

and the proposition is proved.

Finally,

**Definition.** Let $\mathbf{p}, \mathbf{v}, \mathbf{q} \in \mathbf{R}^n$. If $\mathbf{v} \ne \mathbf{0}$, then the *ray from* $\boldsymbol{p}$ *in direction* $\boldsymbol{v}$, denoted by $\text{ray}(\mathbf{p},\mathbf{v})$, is defined by

$$\text{ray}(\mathbf{p},\mathbf{v}) = \{\mathbf{p} + t\mathbf{v} \mid 0 \le t\}.$$

If $\mathbf{p} \ne \mathbf{q}$, then the *ray from* $\boldsymbol{p}$ *through* $\boldsymbol{q}$, denoted by $[\mathbf{pq} >$, is defined by

$$[\mathbf{pq} > = \text{ray}(\mathbf{p},\mathbf{pq}).$$

## 1.3 Angles

The intuitive notion of the angle between two vectors is something that one picks up early, probably while studying Euclidean geometry in high school. In this section we show that there is a very simple rigorous definition of this that is also very easy to compute. Everything we do here holds for an arbitrary real vector space with an inner product, but, for the sake of concreteness, we restrict the discussion to Euclidean space with its standard dot product.

**Definition.** Let $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$. Define the *angle θ between the vectors* ***u*** *and* ***v***, denoted by $\angle(\mathbf{u},\mathbf{v})$, as follows: If either **u** or **v** is the zero vector, then θ is zero; otherwise, θ is that real number such that

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{|\mathbf{u}|\,|\mathbf{v}|}, \quad \text{and} \quad 0 \le \theta \le \pi.$$

Note the purely formal aspect of this definition and that we need the Cauchy-Schwarz inequality to insure that the absolute value of the quotient in (a) is not bigger than 1 (otherwise there would be no such angle). The motivation behind the definition is the law of cosines from Euclidean geometry shown in Figure 1.1. To see this, substitute $|\mathbf{p}|$, $|\mathbf{q}|$, and $|\mathbf{p} + \mathbf{q}|$ for a, b, and c, respectively, and simplify the result.

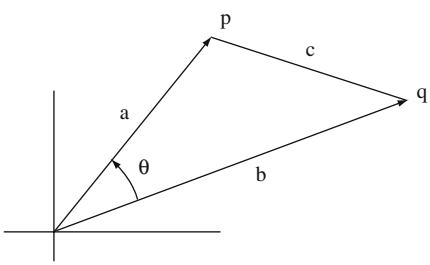
Now if $|\mathbf{u}| = 1$, then

$$\mathbf{u} \bullet \mathbf{v} = |\mathbf{v}| \cos\theta,$$

which one will recognize as the length of the base of the right triangle with hypotenuse **v** and base in the direction of **u**. See Figure 1.2. This means that we can give the following useful interpretation of the dot product:

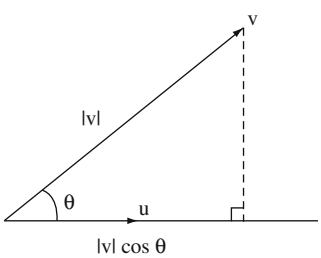
$\mathbf{u} \bullet \mathbf{v}$ is the (signed) length of "the orthogonal projection of **v** on **n**" whenever $|\mathbf{u}| = 1$.

**Definition.** Let $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$. If the angle between the two vectors **u** and **v** is π/2, then they are said to be *perpendicular* and we shall write $\mathbf{u} \perp \mathbf{v}$. If the angle between them is 0 or π, they are said to be *parallel* and we shall write $\mathbf{u} \parallel \mathbf{v}$.




$a^2 + b^2 - 2ab\cos\theta = c^2$

**Figure 1.1.** The law of cosines.




$|u| = 1 \implies uv = |v| \cos\ \theta$

**Figure 1.2.** Interpreting the dot product.

**Definition.** Two vectors **u** and **v** in an **arbitrary** vector space with inner product • are said to be *orthogonal* if $\mathbf{u} \bullet \mathbf{v} = 0$.

**1.3.1. Theorem.** Let $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$.

(1) $\mathbf{u} \perp \mathbf{v}$ if and only if **u** and **v** are orthogonal.
(2) $\mathbf{u} \parallel \mathbf{v}$ if and only if **u** and **v** are linearly dependent.

**Proof.** Most of the theorem follows easily from the definitions. Use the Cauchy-Schwarz inequality to prove (2).

Although the words "orthogonal" and "perpendicular" have different connotations, Theorem 1.3.1 shows that they mean the same thing and we have an extremely easy test for this property, namely, we only need to check that a dot product is zero. Checking whether two vectors are parallel is slightly more complicated. We must check if one is a multiple of the other.

Finally, note that if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is a unit vector, then $u_i = \mathbf{u} \bullet \mathbf{e}_i = \cos\theta_i$, where $\theta_i$ is the angle between **u** and $\mathbf{e}_i$. This justifies the following terminology:

**Definition.** If **v** is a nonzero vector, then the ith component of the unit vector $\frac{1}{|\mathbf{v}|}\mathbf{v}$ is called the *ith direction cosine* of **v**.

## 1.4 Inner Product Spaces: Orthonormal Bases

This section deals with some very important concepts associated with **arbitrary** vector spaces with an inner product. We shall use the dot notation for the inner product. The reader may, for the sake of concreteness, mentally replace every phrase "vector space" with the phrase "vector subspace of $\mathbf{R}^n$ or $\mathbf{C}^n$," but should realize that everything we do here holds in the general setting.

Probably the single most important aspect of inner product spaces is the existence of a particularly nice type of basis.

**Definition.** If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are vectors in an inner product space, we say that they are *mutually orthogonal* if $\mathbf{v}_i \bullet \mathbf{v}_j = 0$ for $i \neq j$. A set of vectors is said to be a *mutually orthogonal set* if it is empty or its vectors are mutually orthogonal.

**Definition.** Let **V** be an inner product space and let B be a basis for **V**. If B is a mutually orthogonal set of vectors, then B is called an *orthogonal basis* for **V**. If, in addition, the vectors of B are all unit vectors, then B is called an *orthonormal basis*. In the special case where **V** consists of only the zero vector, it is convenient to call the empty set an *orthonormal basis for* **V**.

Orthonormal bases are often very useful because they can greatly simplify computations. For example, if we wanted to express a vector **v** in terms of a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then we would normally have to solve the linear equations

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n$$

for the coefficients $a_i$. On the other hand, if we have an orthonormal basis, then it is easy to check that $a_i = \mathbf{v} \bullet \mathbf{v}_i$ and there is nothing to solve. Our first order of business therefore is to describe an algorithm, the *Gram-Schmidt algorithm*, which converts an arbitrary basis into an orthonormal one.

The Gram-Schmidt algorithm is an algorithm that actually can be applied to any collection of vectors and will produce an orthonormal basis for the space spanned by these vectors. We shall illustrate how this process works in the case of two and three vectors.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two nonzero vectors. Then $\mathbf{u}_1 = (1/|\mathbf{v}_1|)\mathbf{v}_1$ is a unit vector. We want to find a unit vector $\mathbf{u}_2$ that is orthogonal to $\mathbf{u}_1$ and so that $\mathbf{u}_1$ and $\mathbf{u}_2$ span the same space as $\mathbf{v}_1$ and $\mathbf{v}_2$. Consider Figure 1.3. If we could find the orthogonal vector $\mathbf{w}$, then all we would have to do is make $\mathbf{w}$ have unit length to get $\mathbf{u}_2$ (assuming that $\mathbf{w}$ is not zero). But $\mathbf{w}$ can easily be computed from the "orthogonal projection" $\mathbf{v}$ of $\mathbf{v}_2$ on $\mathbf{u}_1$ and we pointed out in Section 1.3 that $\mathbf{v}$ could be found using the dot product. The following equations now summarize how one can compute the orthonormal basis $\mathbf{u}_1$ and $\mathbf{u}_2$:

$$\begin{aligned} \mathbf{u}_1 &= \frac{1}{|\mathbf{v}_1|}\mathbf{v}_1 \\ \mathbf{u}_2 &= \frac{1}{|\mathbf{w}|}\mathbf{w} \end{aligned} \tag{1.10}$$

where

$$\begin{aligned} \mathbf{w} &= \mathbf{v}_2 - \mathbf{v} \\ \mathbf{v} &= (\mathbf{v}_2 \bullet \mathbf{u}_1)\mathbf{u}_1. \end{aligned}$$

To see that these computations really do produce orthogonal vectors it suffices to show that the dot product of the vectors $\mathbf{w}$ and $\mathbf{u}_1$ is zero. But

$$\begin{aligned} \mathbf{w} \bullet \mathbf{u}_1 &= [\mathbf{v}_2 - (\mathbf{v}_2 \bullet \mathbf{u}_1)\mathbf{u}_1] \bullet \mathbf{u}_1 \\ &= \mathbf{v}_2 \bullet \mathbf{u}_1 - (\mathbf{v}_2 \bullet \mathbf{u}_1)(\mathbf{u}_1 \bullet \mathbf{u}_1) \\ &= 0. \end{aligned}$$

Next, suppose that we want to construct an orthonormal basis for the space spanned by three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. See Figure 1.4(a). First, apply the construc-

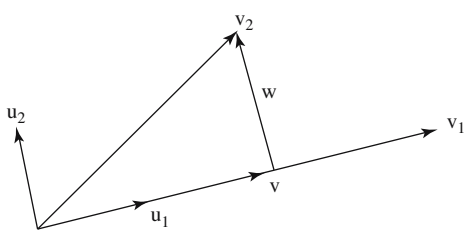



**Figure 1.3.** A simple orthogonal projection.

tion above to find an orthonormal basis for the space spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. Assume that $\mathbf{u}_1$ and $\mathbf{u}_2$ form such a basis. Now find the third vector $\mathbf{u}_3$ by projecting $\mathbf{v}_3$ to the vector $\mathbf{x}$ in the subspace $\mathbf{X}$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. The difference $\mathbf{w} = \mathbf{v}_3 - \mathbf{x}$ is a vector orthogonal to $\mathbf{X}$ that is then normalized to have unit length (assuming again that it is not zero). This leaves one question unanswered, namely, how does one compute $\mathbf{x}$? The example in Figure 1.4(b) motivates the answer. We see that the projection of (1,2,3) onto the plane is (1,2,0). This vector is the sum of two vectors (1,0,0) and (0,2,0), which happen to be the orthogonal projections of (1,2,3) onto the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively. It turns out that the only important property of $\mathbf{e}_1$ and $\mathbf{e}_2$ is that these vectors form an orthonormal basis for the plane. We have now sketched the key ideas needed for the general case. This leads to the recursive construction described in Algorithm 1.4.1.

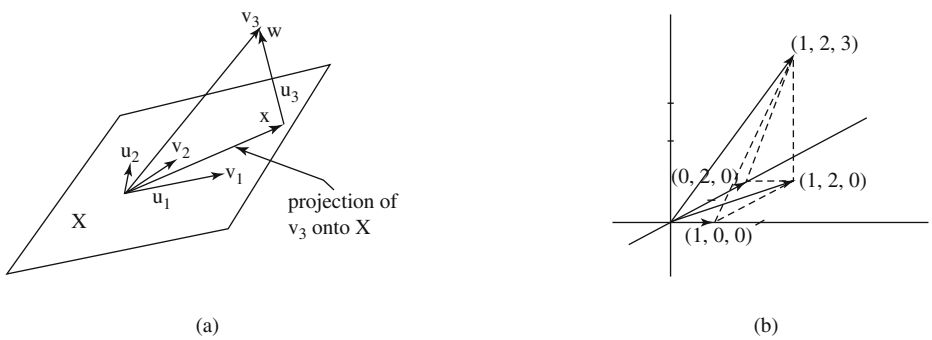



**Figure 1.4.** More orthogonal projections.

**Input:** a set of vectors $S = \{ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \}$
**Output:** an orthonormal basis $B = \{ \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \}$ for span(S)

If $S = \boldsymbol{\phi}$, then return $\boldsymbol{\phi}$.

Let $s = 1$, $B = \boldsymbol{\phi}$, and $m = 0$.

Step 1: If $s > k$, then return B.
Step 2: Let

$$\mathbf{w} = \mathbf{v}_s - (\mathbf{v}_s \bullet \mathbf{u}_1)\,\mathbf{u}_1 - (\mathbf{v}_s \bullet \mathbf{u}_2)\,\mathbf{u}_2 - \ldots - (\mathbf{v}_s \bullet \mathbf{u}_m)\,\mathbf{u}_m.$$

If $\mathbf{w} \neq \mathbf{0}$, then add $\mathbf{u}_{m+1} = (1/|\mathbf{w}|)\,\mathbf{w}$ to B and increment m.
Increment s.
Go to Step 1.

**Algorithm 1.4.1.** The Gram-Schmidt algorithm.

**1.4.1. Theorem.** The Gram-Schmidt algorithm gives the correct result.

**Proof.** There are two parts to proving that the algorithm works. We have to show

(1) the vectors $\mathbf{u}_i$ form an orthonormal set and
(2) they span the same space as the $\mathbf{v}_j$.

One uses induction in both cases. To prove (1) it suffices to check that $\mathbf{w} \bullet \mathbf{u}_i = 0$, $i = 1, 2, \ldots, m$, which is straightforward. This shows that orthogonality is preserved as we go along.

To prove (2), assume inductively that at the beginning of Step 2

$$\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{s-1}) = \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m). \tag{1.11}$$

The inductive hypothesis (1.11) implies that $\mathbf{w}$ belongs to span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s)$, and therefore so does $\mathbf{u}_{m+1}$. This and (1.11) shows that

$$\operatorname{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m+1}) \subseteq \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s). \tag{1.12}$$

Now solve the equation for $\mathbf{w}$ in Step 2 of the algorithm for $\mathbf{v}_s$. Using the inductive hypothesis (1.11), we see that $\mathbf{v}_s$ lies in $\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{s-1}, \mathbf{u}_m)$ and this and another use of the inductive hypothesis (1.11) shows that

$$\operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s) \subseteq \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m+1}). \tag{1.13}$$

The inclusions (1.12) and (1.13) imply that we actually have an equality of sets, proving (2) and the theorem (modulo some special cases such as $\mathbf{w} = \mathbf{0}$ that we leave to the reader).

It should be clear that $m = k$ in the Gram-Schmidt algorithm if and only if the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent. In the worst case, where S is empty or $\mathbf{v}_1 = \mathbf{v}_2 = \ldots = \mathbf{v}_k = \mathbf{0}$, then $m = 0$.

**1.4.2. Corollary.** Every subspace of an inner product space has an orthonormal basis.

**1.4.3. Example.** To find an orthonormal basis $\mathbf{u}_1$ and $\mathbf{u}_2$ for the subspace $\mathbf{X}$ in $\mathbf{R}^3$ spanned by the vectors $\mathbf{v}_1 = (2,-1,1)$ and $\mathbf{v}_2 = (-1,4,0)$.

**Solution.** Applying the Gram-Schmidt algorithm we get

$$\mathbf{u}_1 = \frac{1}{|\mathbf{v}_1|}\mathbf{v}_1 = \frac{1}{\sqrt{6}}(2,-1,1).$$

To get $\mathbf{u}_2$, let

$$\mathbf{v} = (\mathbf{v}_2 \bullet \mathbf{u}_1)\mathbf{u}_1 = (-2,1,-1), \quad \text{and}$$
$$\mathbf{w} = \mathbf{v}_2 - \mathbf{v} = (1,3,1).$$

Then

$$\mathbf{u}_2 = \frac{1}{|\mathbf{w}|}\mathbf{w} = \frac{1}{\sqrt{11}}(1,3,1).$$

One can easily check that $\mathbf{u}_1$ and $\mathbf{u}_2$ are what we want.

**Definition.** Let $\mathbf{X}$ be a subspace of an inner product space $\mathbf{V}$. The *orthogonal complement* of $\mathbf{X}$ in $\mathbf{V}$, denoted by $\mathbf{X}^\perp$, is defined by

$$\mathbf{X}^\perp = \{\mathbf{v} \in \mathbf{V} \mid \mathbf{v} \bullet \mathbf{w} = 0 \text{ for all } \mathbf{w} \in \mathbf{X}\}.$$

Every vector in $\mathbf{X}^\perp$ is called a *normal vector* for $\mathbf{X}$.

**1.4.4. Theorem.** If $\mathbf{X}$ is a subspace of an inner product space $\mathbf{V}$, then the orthogonal complement $\mathbf{X}^\perp$ of $\mathbf{X}$ is a subspace of $\mathbf{V}$ and

$$\mathbf{V} = \mathbf{X} \oplus \mathbf{X}^\perp.$$

Conversely, if

$$\mathbf{V} = \mathbf{X} \oplus \mathbf{Y},$$

where $\mathbf{Y}$ is a subspace with the property that every vector in $\mathbf{Y}$ is normal to $\mathbf{X}$, then $\mathbf{Y} = \mathbf{X}^\perp$.

**Proof.** It is an easy exercise, left to the reader, to show that $\mathbf{X}^\perp$ is a subspace. We prove that $\mathbf{V}$ is a direct sum of $\mathbf{X}$ and $\mathbf{X}^\perp$. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ be an orthonormal basis for $\mathbf{X}$. Define a linear transformation $T : \mathbf{V} \to \mathbf{V}$ by

$$T(\mathbf{v}) = (\mathbf{v} \bullet \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v} \bullet \mathbf{u}_2)\mathbf{u}_2 + \ldots + (\mathbf{v} \bullet \mathbf{u}_k)\mathbf{u}_k \quad (\text{or } \mathbf{0} \text{ if } k = 0).$$

It is easy to check that $\ker(T) = \mathbf{X}^\perp$ and that $\mathbf{v} - T(\mathbf{v})$ belongs to $\ker(T)$. We also have that

$$\mathbf{v} = T(\mathbf{v}) + (\mathbf{v} - T(\mathbf{v})).$$

These facts imply the first part of the theorem. We leave the reader to fill in the details and to prove the converse part (Exercise 1.4.1).

**Definition.** An inner product space $\mathbf{V}$ is said to be the *orthogonal direct sum* of two subspaces $\mathbf{X}$ and $\mathbf{Y}$ if it is a direct sum of $\mathbf{X}$ and $\mathbf{Y}$ and if every vector of $\mathbf{X}$ is orthogonal to every vector of $\mathbf{Y}$.

By Theorem 1.4.4, if $\mathbf{V}$ is an orthogonal direct sum of $\mathbf{X}$ and $\mathbf{Y}$, then $\mathbf{Y} = \mathbf{X}^\perp$. Another consequence of Theorem 1.4.4 is that subspaces can be defined implicitly.

**1.4.5. Theorem.** If $\mathbf{X}$ is a k-dimensional subspace of an n-dimensional inner product space $\mathbf{V}$, then there exist n–k orthonormal vectors $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_{n-k}$, so that

$$\mathbf{X} = \{\mathbf{u} \in \mathbf{V} \mid \mathbf{n}_i \bullet \mathbf{u} = 0 \quad \text{for} \quad 1 \le i \le n-k\}.$$

**Proof.** Choose the vectors $\mathbf{n}_i$ to be an orthonormal basis for the orthogonal complement of $\mathbf{X}$.

Now let $\mathbf{X}$ be a subspace of an inner product space $\mathbf{V}$. Let $\mathbf{v} \in \mathbf{V}$. Since $\mathbf{V} = \mathbf{X} \oplus \mathbf{X}^{\perp}$, we can express $\mathbf{v}$ **uniquely** in the form $\mathbf{v} = \mathbf{x} \oplus \mathbf{y}$, where $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{X}^{\perp}$.

**Definition.** The vector $\mathbf{x}$, denoted by $\mathbf{v}^{\parallel}$, is called the *orthogonal projection of* $\boldsymbol{v}$ *on* $\boldsymbol{X}$ and the vector $\mathbf{y}$, denoted by $\mathbf{v}^{\perp}$, is called the *orthogonal complement of* $\boldsymbol{v}$ *with respect to* $\boldsymbol{X}$.

Note that in the definition, because of the symmetry of the direct sum operator, the orthogonal complement $\mathbf{v}^{\perp}$ of $\mathbf{v}$ with respect to $\mathbf{X}$ is also the orthogonal projection of $\mathbf{v}$ on $\mathbf{X}^{\perp}$. The next theorem shows us how to compute orthogonal projections and complements.

**1.4.6. Theorem.** Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$, $k \ge 1$, be **any** orthonormal basis for a subspace $\mathbf{X}$ in an inner product space $\mathbf{V}$. Let $\mathbf{v} \in \mathbf{V}$. If $\mathbf{v}^{\parallel}$ and $\mathbf{v}^{\perp}$ are the orthogonal projection and orthogonal complement of $\mathbf{v}$ on $\mathbf{X}$, respectively, then

$$\mathbf{v}^{\parallel} = (\mathbf{v} \bullet \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v} \bullet \mathbf{u}_2)\mathbf{u}_2 + \ldots + (\mathbf{v} \bullet \mathbf{u}_k)\mathbf{u}_k \tag{1.14}$$

and

$$\mathbf{v}^{\perp} = \mathbf{v} - (\mathbf{v} \bullet \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{v} \bullet \mathbf{u}_2)\mathbf{u}_2 - \ldots - (\mathbf{v} \bullet \mathbf{u}_k)\mathbf{u}_k. \tag{1.15}$$

**Proof.** Exercise 1.4.2.

In Theorem 1.4.6 it is essential that we have an orthonormal basis, otherwise it is easy to come up with examples that show equations (1.14) and (1.15) are false.

The next definition formalizes some common terminology.

**Definition.** Let $\mathbf{u} \ne \mathbf{0}$ and $\mathbf{v}$ be vectors in an inner product space. Then the *orthogonal projection of* $\boldsymbol{v}$ *on* $\boldsymbol{u}$, denoted by $\mathbf{v}^{\parallel}$, and the *orthogonal complement of* $\boldsymbol{v}$ *with respect to* $\boldsymbol{u}$, denoted by $\mathbf{v}^{\perp}$, are defined by

$$\mathbf{v}^{\parallel} = \left(\mathbf{v} \bullet \frac{\mathbf{u}}{|\mathbf{u}|}\right)\frac{\mathbf{u}}{|\mathbf{u}|} \tag{1.16}$$

and

$$\mathbf{v}^{\perp} = \mathbf{v} - \left(\mathbf{v} \bullet \frac{\mathbf{u}}{|\mathbf{u}|}\right)\frac{\mathbf{u}}{|\mathbf{u}|}. \tag{1.17}$$

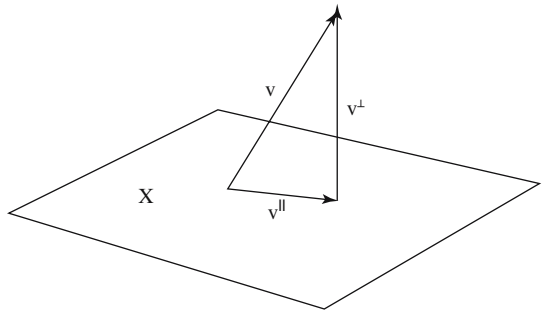



**Figure 1.5.** Decomposing a vector with respect to a subspace.

Clearly, the orthogonal projection of **v** on **u** is the same as the orthogonal projection of **v** on the subspace spanned by **u** and hence is really just a special case of the earlier definition. A similar comment holds for the orthogonal complement. Another way of looking at what we have established is that, given a subspace **X**, every vector **v** can be decomposed into two parts, one "parallel" to **X** and the other orthogonal to it. See Figure 1.5.

We finish this section with a look at some very important classes of matrices.

**Definition.** An $n \times n$ real matrix A is said to be *orthogonal* if $AA^T = A^TA = I$, that is, the inverse of the matrix is just its transpose.

**1.4.7. Lemma.**

(1) The transpose of an orthogonal matrix is an orthogonal matrix.
(2) Orthogonal matrices form a group under matrix multiplication.
(3) The determinant of an orthogonal matrix is $\pm 1$.
(4) The set of orthogonal matrices with determinant +1 forms a subgroup of the group of orthogonal matrices.

**Proof.** Easy.

**Definition.** The group of nonsingular real $n \times n$ matrices under matrix multiplication is called the (real) *linear group* and is denoted by **GL**(n,**R**). The subgroup of orthogonal $n \times n$ matrices is called the *orthogonal group* and is denoted by **O**(n). An orthogonal matrix that has determinant +1 is called a *special orthogonal matrix*. The subgroup of **O**(n) of special orthogonal $n \times n$ matrices is called the *special orthogonal group* and is denoted by **SO**(n).

The groups **SO**(n) and **O**(n) play an important role in many areas of mathematics and much is known about them and their structure. Here are two useful characterizations of orthogonal matrices.

**1.4.8. Theorem.** There is a one-to-one correspondence between orthogonal matrices and orthonormal bases.

**Proof.** If we think of the rows of the matrix as vectors, then we get the correspondence by associating to each matrix the basis of $\mathbf{R}^n$, which consists of the rows of the matrix. A similar correspondence is obtained by using the columns of the matrix.

**1.4.9. Theorem.** Assume $n \geq 1$. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be orthonormal bases in a real inner product space $\mathbf{V}$. If

$$\mathbf{v}_i = \sum_{j=1}^{n} a_{ij}\mathbf{u}_j, \quad a_{ij} \in \mathbf{R}, \tag{1.18}$$

then $A = (a_{ij})$ is an orthogonal matrix. Conversely, let $A = (a_{ij})$ be an orthogonal matrix. If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is an orthonormal basis and if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are defined by equation (1.18), then the $\mathbf{v}$'s will also be an orthonormal basis.

**Proof.** The theorem follows from the following identities

$$\delta_{st} = \mathbf{v}_s \bullet \mathbf{v}_t = \left(\sum_{j=1}^{n} a_{sj}\mathbf{u}_j\right) \bullet \left(\sum_{j=1}^{n} a_{tj}\mathbf{u}_j\right) = \sum_{j=1}^{n} a_{sj}a_{tj}.$$

There is a complex analog of an orthogonal real matrix.

**Definition.** An $n \times n$ complex matrix A is said to be *unitary* if $\overline{A}A^T = A^T\overline{A} = I$, that is, the inverse of the matrix is just its conjugate transpose.

Lemma 1.4.7 remains true if we replace the word "orthogonal" with the word "unitary." In particular, the unitary matrices form a group like the orthogonal ones.

**Definition.** The group of nonsingular complex $n \times n$ matrices under matrix multiplication is called the (complex) *linear group* and is denoted by $\mathbf{GL}(n,\mathbf{C})$. The subgroup of unitary $n \times n$ matrices is called the *unitary group* and is denoted by $\mathbf{U}(n)$. A unitary matrix that has determinant +1 is called a *special unitary matrix*. The subgroup of $\mathbf{U}(n)$ of special unitary $n \times n$ matrices is called the *special unitary group* and is denoted by $\mathbf{SU}(n)$.

The analogs of Theorems 1.4.8 and 1.4.9 hold in the complex case. We omit the details. See, for example, [Lips68] or [NobD77]. We shall run into orthogonal and unitary matrices again later in this chapter and in Chapter 2 when we talk about distance preserving maps or isometries.

## 1.5 Planes

Next, we define the higher-dimensional linear subspaces of Euclidean space. Certainly vector subspaces of $\mathbf{R}^n$ should be such spaces, but "translations" of those should count also.

**Definition.** Any subset $\mathbf{X}$ of $\mathbf{R}^n$ of the form

$$\mathbf{X} = \{\mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \ldots + t_k\mathbf{v}_k \mid t_1, t_2, \ldots, t_k \in \mathbf{R}\}, \tag{1.19a}$$

where $\mathbf{p}$ is a fixed point and the $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are fixed linearly independent vectors in $\mathbf{R}^n$, is called a *k-dimensional plane* (*through* $\boldsymbol{p}$). The *dimension*, k, of $\mathbf{X}$ will be denoted by dim $\mathbf{X}$. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are called a *basis* for the plane.

Clearly, an alternative definition of a k-dimensional plane through a point $\mathbf{p}$ would be to say that it is any set $\mathbf{X}$ of the form

$$\mathbf{X} = \{\mathbf{p} + \mathbf{v} \mid \mathbf{v} \in \mathbf{V}\}, \tag{1.19b}$$

where $\mathbf{V}$ is a k-dimensional vector subspace of $\mathbf{R}^n$. Furthermore, the subspace $\mathbf{V}$ is uniquely determined by $\mathbf{X}$ (Exercise 1.5.1).

The (n – 1)-dimensional planes in $\mathbf{R}^n$ are especially interesting.

**Definition.** Any subset $\mathbf{X}$ of $\mathbf{R}^n$ of the form

$$\{\mathbf{p} \mid \mathbf{n} \bullet \mathbf{p} = d\}, \tag{1.20}$$

where $\mathbf{n}$ is a fixed **nonzero** vector of $\mathbf{R}^n$ and d is a fixed real number, is called a *hyperplane*.

Note that if $\mathbf{n} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{p} = (x_1, x_2, \ldots, x_n)$, then the equation in (1.20) is equivalent to the usual form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = d \tag{1.21}$$

of the equation for a hyperplane. Note also that if $\mathbf{p}_0$ belongs to the hyperplane, then by definition $d = \mathbf{n} \bullet \mathbf{p}_0$ and we can rewrite the equation for the hyperplane in the form

$$\mathbf{n} \bullet (\mathbf{p} - \mathbf{p}_0) = 0. \tag{1.22}$$

Equation (1.22) says that the hyperplane $\mathbf{X}$ consists of those points $\mathbf{p}$ with the property that the vector $\mathbf{p} - \mathbf{p}_0$ is orthogonal to the vector $\mathbf{n}$. See Figure 1.6.

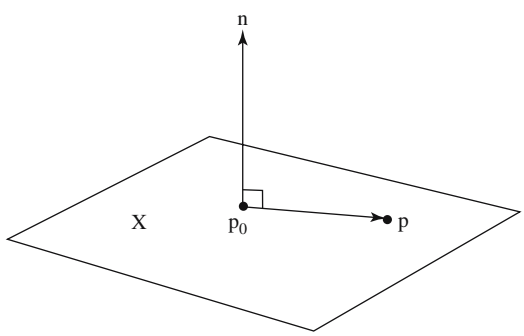



**Figure 1.6.** The point-normal definition of a hyperplane.

**Definition.** Equation (1.22) is called the *point-normal form* of the equation for the hyperplane defined by (1.20) or (1.21). The vector **n** is called a *normal* vector to the hyperplane.

**1.5.1. Example.** Consider the hyperplane defined by z = 0. This equation can be rewritten in the form

$$0(x-0)+0(y-0)+1(z-0)=0.$$

Note that (0,0,0) is a point in the hyperplane and (0,0,1) is a normal vector for it.

The next proposition justifies the phrase "plane" in the word "hyperplane."

**1.5.2. Proposition.**

(1) A hyperplane **X** in $\mathbf{R}^n$ is an (n – 1)-dimensional plane. If **X** is defined by the equation $\mathbf{n} \bullet \mathbf{p} = d$, then any basis for the vector subspace

$$\mathbf{K} = \{\mathbf{p} \in \mathbf{R}^n \mid \mathbf{n} \bullet \mathbf{p} = 0\}$$

is a basis for **X**.

(2) Conversely, every (n – 1)-dimensional plane in $\mathbf{R}^n$ is a hyperplane.

**Proof.** To prove (1) note first that **K** is a vector subspace. This can be seen either by a direct proof or by observing that **K** is the kernel of the linear transformation

$$T: \mathbf{R}^n \to \mathbf{R}$$

defined by

$$T(\mathbf{p}) = \mathbf{n} \bullet \mathbf{p}.$$

It follows easily from Theorem B.10.3 that **K** is an (n – 1)-dimensional vector subspace of $\mathbf{R}^n$. If $\mathbf{p}_0$ is any point of **X**, then it is easy to show that

$$\mathbf{X} = \{\mathbf{p}_0 + \mathbf{q} \mid \mathbf{q} \in \mathbf{K}\},$$

proving the first part of the lemma. The converse, part (2), follows from Theorem 1.4.5. Exercise 1.5.2 asks the reader to fill in missing details.

**1.5.3. Example.** To find a basis for the (hyper)plane **X** in $\mathbf{R}^3$ defined by 2x + y – 3z = 6.

**Solution.** There will be two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in our basis. We use Proposition 1.5.2(1). The vector $\mathbf{n} = (2,1,-3)$ is a normal vector for our plane. Therefore, to find $\mathbf{v}_1$ and $\mathbf{v}_2$ is to find a basis for the kernel **K** of the map

$$\mathbf{p} \to \mathbf{n} \bullet \mathbf{p}.$$

The direct approach is to solve the equation $\mathbf{n} \bullet \mathbf{p} = 0$, that is, $2x + y - 3z = 0$, for two noncollinear points $\mathbf{v}_1$ and $\mathbf{v}_2$. Alternatively, compute three noncollinear points $\mathbf{p}_0$, $\mathbf{p}_1$, and $\mathbf{p}_2$ in $\mathbf{X}$ and set $\mathbf{v}_1 = \mathbf{p}_0\mathbf{p}_1$ and $\mathbf{v}_2 = \mathbf{p}_0\mathbf{p}_2$. For example, $\mathbf{p}_0 = (1,1,-1)$, $\mathbf{p}_1 = (3,0,0)$, and $\mathbf{p}_2 = (0,6,0)$ would give $\mathbf{v}_1 = (2,-1,1)$ and $\mathbf{v}_2 = (-1,5,0)$. By construction these vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ will also be a basis for $\mathbf{K}$. The first approach that involves solving an equation for only two points rather than solving the equation $2x + y - 3z = 6$ for three points is obviously simpler; however, in other problems a plane may not be defined by an equation.

Example 1.5.3 shows how one can find a basis for a plane if one knows some points in it. A related question in the case of hyperplanes is to find the **equation** for it given some points in it. To answer that question in $\mathbf{R}^3$ one can use the cross product.

**Definition.** Let $\mathbf{v}, \mathbf{w} \in \mathbf{R}^3$. Define the *cross product* $\mathbf{v} \times \mathbf{w} \in \mathbf{R}^3$ by

$$\mathbf{v} \times \mathbf{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1). \tag{1.23}$$

Now, formula (1.23) is rather complicated. The standard trick to make it easier to remember is to take the formal determinant of the following matrix:

$$\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

The coefficients of the symbols $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ will then be the x-, y-, and z-component, respectively, of the cross product.

We shall look at the cross product and its properties more carefully later in Section 1.10. Right now we shall only make use of the fact that the cross product of two vectors produces a vector that is orthogonal to both of these vectors, something easily checked from the formula.

**1.5.4. Example.** To find an equation for the hyperplane that contains the points $\mathbf{p} = (1,0,1)$, $\mathbf{q} = (1,2,0)$, and $\mathbf{r} = (0,0,3)$.

**Solution.** We have that

$$\mathbf{pq} = (0,2,-1), \quad \mathbf{pr} = (-1,0,2), \quad \text{and} \quad \mathbf{pq} \times \mathbf{pr} = (4,1,2).$$

Therefore, an equation for the plane is

$$(4,1,2) \bullet ((x,y,z) - (1,0,1)) = 0,$$

which reduces to

$$4x + y + 2z = 6.$$

If we compare arbitrary k-dimensional planes and hyperplanes, we see that the former have so far only an explicit definition in terms of parameterizations whereas

the latter can also be defined implicitly via an equation using a normal vector. Actually, Theorem 1.4.5 corrects this situation and shows that an arbitrary k-dimensional plane **X** can also be defined by means of normal vectors and hence an equation in the following sense: If $\mathbf{p}_0$ is any point in the plane, then there exist n–k orthonormal vectors $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_{n-k}$, so that

$$\mathbf{X} = \{\mathbf{p} \mid \mathbf{n}_i \bullet (\mathbf{p} - \mathbf{p}_0) = 0 \quad \text{for} \quad 1 \le i \le n - k\}. \tag{1.24}$$

**Definition.** Equation (1.24) is called the *point-normals form* of the equation for the plane **X**.

Now normal vectors to hyperplanes are not unique, because any nonzero multiple will determine the same hyperplane.

**1.5.5. Lemma.** If $\mathbf{n}_1$ and $\mathbf{n}_2$ are two normal vectors for a hyperplane **X**, then $\mathbf{n}_1$ and $\mathbf{n}_2$ are parallel.

**Proof.** By hypothesis, **X** is defined by equations

$$\mathbf{n}_i \bullet \mathbf{p} = d_i.$$

Replacing $\mathbf{n}_2$ by a nonzero multiple if necessary, we may assume that $d_1 = d_2$. Therefore,

$$\mathbf{n}_1 \bullet \mathbf{p} = \mathbf{n}_2 \bullet \mathbf{p}$$

for all **p** in **X**. It follows that

$$\mathbf{n}_1 \bullet \mathbf{p} = 0$$

and

$$\mathbf{n}_2 \bullet \mathbf{p} = 0$$

define the same hyperplane **Y**. But **Y** is an (n – 1)-dimensional vector subspace of $\mathbf{R}^n$ and so has a unique one-dimensional orthogonal complement (Theorem 1.4.4). Since the normal vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ belong to this complement, they must be multiples of each other and the lemma is proved.

Lemma 1.5.5 justifies the following definition:

**Definition.** Two hyperplanes are said to be *parallel* if they have parallel normal vectors. Two hyperplanes are said to be *orthogonal* if they have orthogonal normal vectors. A vector is said to be *parallel* or *orthogonal* to a hyperplane if it is orthogonal or parallel, respectively, to a normal vector of the hyperplane.

Although we shall not do so here (except in the case of "oriented" hyperplanes later on), it is actually possible to define an angle between arbitrary planes. See [IpsM95], for example. One could then define parallel and orthogonal in terms of that angle like we did for vectors. At any rate, with our definition, we are calling any two

hyperplanes defined by equations $\mathbf{n} \bullet \mathbf{p} = d_1$ and $\mathbf{n} \bullet \mathbf{p} = d_2$ parallel. They also have the same bases. It is useful to generalize these definitions.

**Definition.** Let **X** and **Y** be s- and t-dimensional planes, respectively, with $s \le t$. If **Y** has a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_t$, so that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$ is a basis for **X**, then we say that **X** is *parallel* to **Y** and **Y** is *parallel* to **X**.

**1.5.6. Lemma.** In the case of hyperplanes the two notions of parallel agree.

**Proof.** Exercise 1.5.5.

Next, we want to extend the notion of orthogonal projection and orthogonal complement of vectors to planes. Let **X** be a k-dimensional plane with basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. Let $\mathbf{X}_0$ be the vector subspace generated by the vectors $\mathbf{v}_i$, that is,

$$\mathbf{X}_0 = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k).$$

Note that $\mathbf{X}_0$ is a plane through the origin parallel to **X**.

**1.5.7. Lemma.** The plane $\mathbf{X}_0$ is independent of the choice of basis for **X**.

**Proof.** Exercise 1.5.6.

**Definition.** Let **v** be a vector. The *orthogonal projection of **v** on **X*** is the orthogonal projection of **v** on $\mathbf{X}_0$. The *orthogonal complement of **v** with respect to **X*** is the orthogonal complement of **v** with respect to $\mathbf{X}_0$.

By Lemma 1.5.7, the orthogonal projection of a vector on a plane and its orthogonal complement is well defined. We can use Theorem 1.4.6 to compute them.

A related definition is

**Definition.** A vector is said to be *parallel* to a plane if it lies in the subspace spanned by any basis for the plane. A vector is said to be *orthogonal* to a plane if it is orthogonal to all vectors in any basis for the plane. More generally, a plane **X** is said to be *parallel* to a plane **Y** if every vector in a basis for **X** is parallel to **Y** and **X** is *orthogonal* to **Y** if every vector in a basis for **X** is orthogonal to **Y**.

It is easy to show that the notion of a vector or plane being parallel or orthogonal to another plane does not depend on the choice of bases for the planes. Note that, as a special case, a vector will be parallel to a line if and only if it is parallel to any direction vector for the line. Another useful observation generalizes and makes more precise a comment in the last section. Specifically, given an arbitrary plane **X** in $\mathbf{R}^n$, any vector **v** in $\mathbf{R}^n$ can be decomposed into a part that is parallel to **X** and a part that is orthogonal to it. See Figure 1.5 again. Finally, the new notion of parallel and orthogonal planes agrees with the earlier one.

**1.5.8. Example.** To find the equation for the plane **X** in $\mathbf{R}^3$ through the point $\mathbf{p}_0 = (1,3,2)$, which is parallel to the line

$$x = 2 + 3t$$
$$y = -t$$
$$z = 7$$

and orthogonal to the plane $x - z = 2$ .

**Solution.** If $\mathbf{n} = (a,b,c)$ is a normal for $\mathbf{X}$, then $\mathbf{n}$ must be orthogonal to the direction vector $(3,-1,0)$ for the given line and orthogonal to the normal $(1,0,-1)$ for the given plane, that is,

$$3a - b = 0$$

and

$$a - c = 0.$$

Solving these two equations gives that $b = 3a$ and $c = a$. In other words, $(a,3a,a)$ is a normal vector for $\mathbf{X}$. It follows that

$$(1,3,1) \bullet ((x,y,z) - (1,3,2)) = 0$$

or

$$x + 3y + z = 12$$

is an equation for $\mathbf{X}$.

We finish this section with two more definitions. The first generalizes the halfplanes $\mathbf{R}^n_+$ and $\mathbf{R}^n_-$.

**Definition.** Let $\mathbf{p}_0$, $\mathbf{n} \in \mathbf{R}^n$ with $\mathbf{n} \neq \mathbf{0}$. The sets

$$\{\mathbf{p} \in \mathbf{R}^n \mid \mathbf{n} \bullet (\mathbf{p} - \mathbf{p}_0) \geq 0\}$$

and

$$\{\mathbf{p} \in \mathbf{R}^n \mid \mathbf{n} \bullet (\mathbf{p} - \mathbf{p}_0) \leq 0\}$$

are called the *halfplanes* determined by the hyperplane $\mathbf{n} \bullet (\mathbf{p} - \mathbf{p}_0) = 0$. A *halfline* is a halfplane in $\mathbf{R}$.

A hyperplane in $\mathbf{R}^n$ divides $\mathbf{R}^n$ into three parts: itself and the two halfplanes on either "side" of it. Figure 1.7 shows the two halfplanes in the plane defined by the line (hyperplane) $2x + 3y - 6 = 0$.

Sometimes one needs to talk about the smallest plane spanned by a set.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$. The *affine hull* or *affine closure* of $\mathbf{X}$, denoted by aff $(\mathbf{X})$, is defined by

$$\text{aff}(\mathbf{X}) = \cap \{\mathbf{P} \mid \mathbf{P} \text{ is a plane which contains } \mathbf{X}\}.$$

The following lemma justifies the definition of the affine hull of a set:

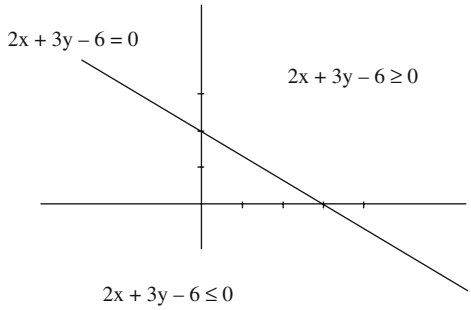



**Figure 1.7.** The halfplanes defined by the line $2x + 3y - 6 = 0$.

**1.5.9. Lemma.**

(1) The intersection of an arbitrary number of planes is a plane.
(2) If $\mathbf{X}$ is a plane, then aff $(\mathbf{X}) = \mathbf{X}$.

**Proof.** This is left as an exercise for the reader (Exercises 1.5.3 and 1.5.4).

It follows from the lemma that affine hulls are actually planes. One can also easily see that aff($\mathbf{X}$) is contained in any plane that contains $\mathbf{X}$, which is why one refers to it as the "smallest" such plane.

**1.5.10. Theorem.** Let $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k \in \mathbf{R}^n$. Then

$$\mathrm{aff}(\{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k\}) = \{\mathbf{p}_0 + t_1\mathbf{p}_0\mathbf{p}_1 + \ldots + t_k\mathbf{p}_0\mathbf{p}_k \mid t_i \in \mathbf{R}\}$$

**Proof.** Exercise 1.5.8.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two planes in $\mathbf{R}^n$. The definition implies that $\mathbf{X}$ and $\mathbf{Y}$ are the translations of **unique** vector subspaces $\mathbf{V}$ and $\mathbf{W}$, respectively, that is,

$$\mathbf{X} = \{\mathbf{p} + \mathbf{v} \mid \mathbf{v} \in \mathbf{V}\} \quad \text{and} \quad \mathbf{Y} = \{\mathbf{q} + \mathbf{w} \mid \mathbf{w} \in \mathbf{W}\}$$

for some $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$.

**Definition.** The planes $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbf{R}^n$ are said to be *transverse* if

$$\dim(\mathbf{V} \cap \mathbf{W}) = \max\{0, \dim(\mathbf{V}) + \dim(\mathbf{W}) - n\}.$$

Two transverse lines in $\mathbf{R}^3$ are said to be *skew*.

Intuitively, two planes are transverse if their associated subspaces $\mathbf{V}$ and $\mathbf{W}$ span as high-dimensional space as possible given their dimensions. To put it another way, the intersection of $\mathbf{V}$ and $\mathbf{W}$ should be as small as possible. Sometimes this is referred to as the planes being in *general position*. For example, the x- and y-axes are transverse in $\mathbf{R}^n$, but the x-axis and the parallel line defined by $y = 1$ are not. The xy- and yz-plane are transverse in $\mathbf{R}^3$ but not in $\mathbf{R}^4$.

## 1.6 Orientation

This section is an introduction to the concept of orientation. Although this intuitive concept is familiar to everyone, probably few people have thought about what it means and how one could give a precise definition.

The notion of orientation manifests itself in many different contexts. In everyday conversation one encounters phrases such as "to the left of," "to the right of," "clockwise," or "counterclockwise." Physicists talk about right- or left-handed coordinate systems. In computer graphics, one may want to pick normals to a planar curve in a consistent way so that they all, say, point "inside" the curve. See Figure 1.8. A similar question might be asked for normals in the case of surfaces. How can one tell in a systematic way that our choice of normals is "consistent"? What does this really mean?

Probably the easiest way to demonstrate the orientability property for surfaces is in terms of the number of "sides" that they have. Consider the cylinder in Figure 1.9(a). This surface has the property that if one were a bug, the only way to get from the "outside" to the "inside" would be to crawl over the edge. We express this by saying that the cylinder is "two-sided" or orientable. Now, a cylinder can be obtained from a strip of paper by gluing the two ends together in the obvious way. If, on the other hand, we take this same strip of paper and first give it a 180-degree twist before we glue the ends together, then we will get what is called a *Moebius strip* (discovered by A.F. Moebius and independently by J.B. Listing in 1858). See Figure 1.9(b). Although

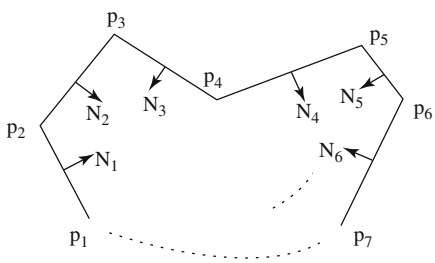



**Figure 1.8.** Uniformly oriented normals.

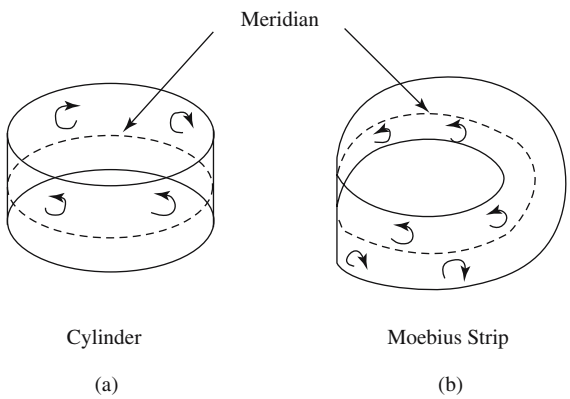



**Figure 1.9.** Induced orientations along paths.

the strip has two sides at any given point, we can get from one side to the other by walking all the way around the strip parallel to the meridian. The Moebius strip is a "one-sided" or nonorientable surface. In general, a simple-minded definition is to say that a surface **S** is orientable (nonorientable) if one cannot (can) get from one side of **S** at a point to the other side by walking along the surface.

One can define orientability also in terms of properties that relate more directly to the intuitive meaning of "orient." For example, an orientable surface is one where it is possible to define a consistent notion of left and right or clockwise and counterclockwise. But what does "consistent" mean? If two persons are standing at different points of a surface and they each have decided what to call clockwise, how can they determine whether their choices are consistent (assuming that they cannot see each other)? One way to answer this question is to have one of them walk over to where the other one is standing and then compare their notions of clockwise. This leads to the following approach to defining a consistent orientation at every point of a surface **S**. Starting at a point **p** on the surface choose an orientation at **p** by deciding which of the two possible rotations around the point is to be called clockwise. Now let **q** be any other point of **S** (**q** may be equal to **p**). Walk to **q** along some path, all the while remembering which rotation had been called clockwise. This will induce a notion of clockwise for rotations at **q**, and hence an orientation at **q**. Unfortunately, there are many paths from **p** to **q** (nor is there a unique shortest path in general) and, although this may not seem immediately obvious, different paths may induce different orientations. If an orientation at **p** always induces the same orientation at every point of the surface no matter which path we take to that point, then **S** is called orientable. Figure 1.9(b) shows that walking around the meridian of the Moebius strip will induce an orientation back at the starting point that is opposite to the one picked at the beginning. Therefore, we would call the Moebius strip nonorientable, and our new definition is compatible with the earlier one.

Orientability is an intrinsic property of surfaces. F. Klein was the first to observe this fact explicitly in 1876. The sphere is orientable, as are the torus (the surface of a doughnut) and double torus (the surface of a solid figure eight) shown in Figure 1.10. Actually, since the torus will be a frequent example, this is a good time to give a slightly more precise definition of it. It is a special case of a more general type of surface.

**Definition.** A *surface of revolution* in $\mathbf{R}^3$ is a space **S** obtained by revolving a planar curve about a line in that plane called the *axis of revolution*. A *meridian* of **S** is a con-

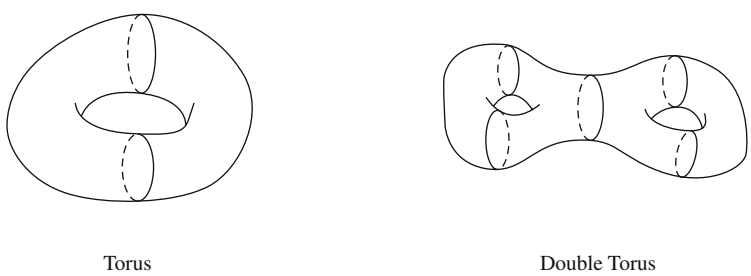



**Figure 1.10.** Orientable surfaces.

nected component of the intersection of **S** and a plane through the axis of revolution. A *circle of latitude* of **S** is a connected component of the intersection of **S** and a plane orthogonal to the axis of revolution. A *torus* is a surface of revolution where the curve being revolved is a circle that does not intersect the axis of revolution.

See Figure 1.11. Note that meridians of surfaces of revolution meet their circles of latitude in a single point. Note also that a surface of revolution may not actually be a "surface" if the curve being revolved is not chosen carefully, for example, if it intersects the axis. (The term "surface" will be defined carefully in Chapter 5.) Surfaces of revolution are also orientable.

There are surfaces **without** boundary that are nonorientable and the reader is challenged to find one on his own (or wait until Chapter 6). One word of caution though: Nonorientable surfaces without boundary cannot be found in $\mathbf{R}^3$ (see Exercise 6.5.1.). One needs a fourth dimension.

Enough of this intuitive discussion of orientability. Let us move on to mathematical definitions. In this section we define the most basic concept, namely, what is meant by the orientation of a vector space. This corresponds to a definition of the local concept, that is, the notion of an orientation at a point.

Consider the problem of trying to define an orientation at the origin of $\mathbf{R}^2$. Let $(\mathbf{v}_1,\mathbf{v}_2)$ be an ordered basis. See Figure 1.12. We could use this ordered pair to suggest the idea of counterclockwise motion. The only trouble is that there are many ordered bases for $\mathbf{R}^2$. For example, the pair $(\mathbf{w}_1,\mathbf{w}_2)$ in Figure 1.12 also corresponds to coun-

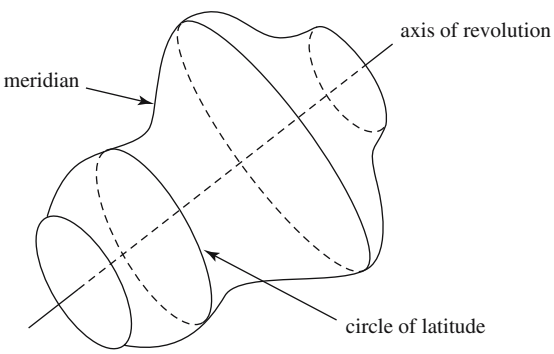



**Figure 1.11.** A surface of revolution.

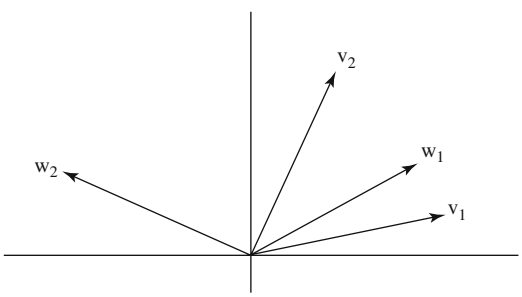



**Figure 1.12.** Using ordered bases to define an orientation.

terclockwise motion. Therefore, we need an appropriate equivalence relation. The key to defining this relation is the matrix relating two ordered bases.

Let $(\mathbf{v}_1,\mathbf{v}_2)$ and $(\mathbf{w}_1,\mathbf{w}_2)$ be two ordered bases. Suppose that

$$\mathbf{v}_i = a_{i1}\mathbf{w}_1 + a_{i2}\mathbf{w}_2,$$

for $a_{ij} \in \mathbf{R}$. Define $(\mathbf{v}_1,\mathbf{v}_2)$ to be equivalent to $(\mathbf{w}_1,\mathbf{w}_2)$ if the determinant of the matrix $(a_{ij})$ is positive. Since we are dealing with bases, we know that the $a_{ij}$ exist and are unique and that the matrix $(a_{ij})$ is nonsingular. It is easy to see that our relation is an equivalence relation and that we have precisely two equivalence classes because the nonzero determinant is either positive or negative. We could define an orientation of $\mathbf{R}^2$ to be such an equivalence class. As a quick check to see that we are getting what we want, note that if $\mathbf{w}_1 = \mathbf{v}_2$ and $\mathbf{w}_2 = \mathbf{v}_1$, then

$$(a_{ij}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the determinant of this matrix is $-1$, so that $(\mathbf{v}_1,\mathbf{v}_2)$ and $(\mathbf{v}_2,\mathbf{v}_1)$ determine different equivalence classes.

Because we only used vector space concepts, it is easy to generalize what we just did.

**Definition.** Let $B_1 = (\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ and $B_2 = (\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_n)$ be ordered bases for a vector space $\mathbf{V}$ and let

$$\mathbf{w}_i = \sum_{j=1}^{n} a_{ij}\mathbf{v}_j, \quad \text{where} \quad a_{ij} \in \mathbf{R}.$$

We say that $B_1$ is *equivalent* to $B_2$, and write $B_1 \sim B_2$ if the determinant of the matrix $(a_{ij})$ is positive.

**1.6.1. Lemma.** $\sim$ is an equivalence relation on the set of ordered bases for $\mathbf{V}$ with precisely two equivalence classes.

**Proof.** Exercise 1.6.1.

**Definition.** An *orientation* of a vector space $\mathbf{V}$ is defined to be an equivalence class of ordered bases of $\mathbf{V}$ with respect to the relation $\sim$. Given one orientation of $\mathbf{V}$, then the other one is called the *opposite orientation*. The equivalence class of an ordered basis $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ will be denoted by $[\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n]$. We shall say that the ordered basis $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ *induces* or *determines* the orientation $[\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n]$. An *oriented vector space* is a pair $(\mathbf{V},\sigma)$, where $\mathbf{V}$ is vector space and $\sigma$ is an orientation of it.

**1.6.2. Example.** To show that the ordered bases $((1,3),(2,1))$ and $((1,1),(2,0))$ determine the same orientation of the plane.

**Solution.** See Figure 1.13. Note that

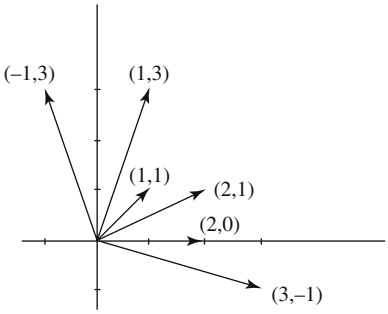



**Figure 1.13.** Ordered bases.

$$(1,1) = \frac{1}{5}(1,3) + \frac{2}{5}(2,1)$$
$$(2,0) = -\frac{2}{5}(1,3) + \frac{6}{5}(2,1)$$

and

$$\det\begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{2}{5} & \frac{6}{5} \end{pmatrix} = 2 > 0.$$

**1.6.3. Example.** To show that the ordered bases ((1,3),(2,1)) and ((3,–1),(–1,3)) determine different orientations of the plane.

**Solution.** See Figure 1.13. Note that

$$(3,-1) = -(1,3) + 2(2,1)$$
$$(-1,3) = \frac{7}{5}(1,3) - \frac{6}{5}(2,1)$$

and

$$\det\begin{pmatrix} -1 & 2 \\ \frac{7}{5} & -\frac{6}{5} \end{pmatrix} = -\frac{8}{5} < 0.$$

Since arbitrary vector spaces do not have any special bases, one typically cannot talk about a "standard" orientation, but can only **compare** ordered bases as to whether they determine the same orientation or not. In the special case of $\mathbf{R}^n$ we do have the standard basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ though.

**Definition.** $[\mathbf{e}_1,\mathbf{e}_2, \ldots ,\mathbf{e}_n]$ is called the *standard orientation* of $\mathbf{R}^n$.

The standard orientation corresponds to what is called a *right-handed coordinate system* and the opposite orientation to a *left-handed coordinate system*.

It should be pointed out that the really important concept here is not the formal definition of an orientation but rather the associated terminology. It is phrases like "these two ordered bases determine the same or opposite orientations" or "this basis induces the standard or non-standard orientation of $\mathbf{R}^n$" that the reader needs to understand.

Solving linear equations can be tedious and therefore it is nice to know that there is a much simpler method for determining whether or not ordered bases determine the same orientation or not in the case of $\mathbf{R}^n$.

**1.6.4. Lemma.** Two ordered bases $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ and $(\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_n)$ of $\mathbf{R}^n$ determine the same orientation if and only if

$$\det\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{n-1} \\ \mathbf{v}_n \end{pmatrix} \quad \text{and} \quad \det\begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{n-1} \\ \mathbf{w}_n \end{pmatrix}$$

have the same sign.

**Proof.** The details of the proof are left to the reader. The idea is to relate both bases to the standard ordered basis $(\mathbf{e}_1,\mathbf{e}_2, \ldots ,\mathbf{e}_n)$.

**1.6.5. Example.** The solutions to Examples 1.6.2 and 1.6.3 above are much easier using Lemma 1.6.4. One does not have to solve any linear equations but simply has to compute the following determinants:

$$\det\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} = -5 \quad \det\begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} = -2 \quad \det\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} = 8$$

**Definition.** Let $\mathbf{V}$ be a vector space. A nonsingular linear transformation $T : \mathbf{V} \to \mathbf{V}$ is said to be *orientation preserving* (or *sense preserving*) if $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ and $(T(\mathbf{v}_1),T(\mathbf{v}_2), \ldots ,T(\mathbf{v}_n))$ determine the same orientation of $\mathbf{V}$ for all ordered bases $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ of $\mathbf{V}$. If T is not orientation preserving then it is said to be *orientation reversing* (or *sense reversing*). More generally, if $(\mathbf{V},\sigma)$ and $(\mathbf{W},\tau)$ are two oriented n-dimensional vector spaces and if $T : \mathbf{V} \to \mathbf{W}$ is a nonsingular linear transformation (that is, an isomorphism), then T is said to be *orientation preserving* if $\tau = [T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots ,T(\mathbf{v}_n)]$ for all ordered bases $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ of $\mathbf{V}$ with the property that $\sigma = [\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n]$; otherwise, T is said to be *orientation reversing*.

The identity map for a vector space is clearly orientation preserving. Exercise 1.6.7 asks you to show that whether or not a map is orientation preserving or reversing can

be determined by checking the property on a **single** ordered basis. In the case of an arbitrary linear transformation from a vector space to itself there is another simple test for when it is orientation preserving or reversing.

**1.6.6. Theorem.** Let $\mathbf{V}$ be a vector space and let $T : \mathbf{V} \to \mathbf{V}$ be a nonsingular linear transformation. The transformation T is orientation preserving if and only if $\det(T) > 0$.

**Proof.** This theorem follows immediately from the definitions that are involved.

**1.6.7. Theorem.** Let $\mathbf{V}$ be a vector space and let $T, T_i : \mathbf{V} \to \mathbf{V}$ be nonsingular linear transformations.

(1) The transformation T is orientation preserving if and only if $T^{-1}$ is.
(2) Let $T = T_1 \circ T_2 \circ \cdots \circ T_k : \mathbf{V} \to \mathbf{V}$. The transformation T is orientation preserving if and only if the number of transformations $T_i$ that are orientation reversing is even.

**Proof.** This theorem is an immediate consequence of Theorem 1.6.6 and the identities

$$\det(T^{-1}) = \frac{1}{\det(T)} \quad \text{and} \quad \det(T) = \det(T_1)\det(T_2)\cdots\det(T_k).$$

**Definition.** Let $\mathbf{X}$ be a plane in $\mathbf{R}^n$ with basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. An *orientation* of $\mathbf{X}$ is an orientation of the linear subspace $\mathrm{aff}(\{\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_k\})$ (which is $\mathbf{X}$ translated to the origin) of $\mathbf{R}^n$. An *oriented plane* is a pair $(\mathbf{X},\sigma)$, where $\mathbf{X}$ is a plane and $\sigma$ is an orientation of $\mathbf{X}$. The expression "the plane $\mathbf{X}$ *oriented by* (the ordered basis) $(\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_k)$" will mean the oriented plane $(\mathbf{X}, [\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_k])$. An oriented line is often called a *directed line*.

An oriented plane $(\mathbf{X},\sigma)$ will often be referred to simply as the "oriented plane $\mathbf{X}$." In that case the orientation $\sigma$ is assumed given but just not stated explicitly until it is needed. The orientation of an oriented line is defined by a **unique** unit direction vector.

Normally, although they seem to make sense, expressions such as "the angle between two lines" or "the angle between two planes in $\mathbf{R}^3$" are ambiguous because it could mean one of two angles. In the oriented case one can make sense of that however.

**Definition.** Let $(\mathbf{X},\sigma)$ and $(\mathbf{Y},\tau)$ be oriented hyperplanes in $\mathbf{R}^n$. Let $\sigma = [\mathbf{v}_1,\mathbf{v}_2, \ldots , \mathbf{v}_{n-1}]$ and $\tau = [\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_{n-1}]$. If $\mathbf{v}_n$ and $\mathbf{w}_n$ are normal vectors for $\mathbf{X}$ and $\mathbf{Y}$, respectively, with the property that $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ and $(\mathbf{w}_1,\mathbf{w}_2, \ldots ,\mathbf{w}_n)$ induce the standard orientation of $\mathbf{R}^n$, then the angle between the vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ is called the *angle between the oriented hyperplanes* $(\mathbf{X},\sigma)$ and $(\mathbf{Y},\tau)$.

The angle between oriented hyperplanes is well defined (Exercise 1.6.5).

**Definition.** Let **L** be an oriented line and let **u** be the unit vector that defines the orientation of **L**. Let **p** and **q** be two points on **L**. The *oriented* or *signed distance from* ***p*** *to* ***q***, denoted by $\|\mathbf{pq}\|$, is defined by

$$\|\mathbf{pq}\| = \mathbf{pq} \bullet \mathbf{u}.$$

It is easy to check that if $\mathbf{p} \neq \mathbf{q}$, then $\|\mathbf{pq}\|$ is just the ordinary (unsigned) distance $|\mathbf{pq}|$ if the vector **pq** induces the same orientation on **L** as **u** and $-|\mathbf{pq}|$ otherwise (Exercise 1.6.6).

The angle between two vectors as defined in Section 1.3 is always a nonnegative quantity, but sometimes it is convenient to talk about a signed angle, where the sign of the angle is determined by the direction (counterclockwise or clockwise) that the angle "sweeps" out.

**Definition.** Let **u** and **v** be two linearly independent vectors in the plane $\mathbf{R}^2$. If $\theta$ is the angle between **u** and **v**, define $\angle_s(\mathbf{u},\mathbf{v})$, the *signed angle between* ***u*** *and* ***v***, by

$$\begin{aligned} \angle_s(\mathbf{u},\mathbf{v}) &= \theta, && \text{if the ordered basis } (\mathbf{u},\mathbf{v}) \text{ induces the standard orientation of } \mathbf{R}^2 \\ &= -\theta, && \text{otherwise.} \end{aligned}$$

This finishes our discussion of the local theory of orientation. We shall return to the subject of orientation in Chapters 6 and 8 and define what is meant by an orientation at a point of a "curved" space. We shall also consider global aspects of orientation and what it might mean to say that an entire space is oriented. However, in order not to leave the reader in a kind of limbo with respect to how the definitions of this section fit into the whole picture, it is useful to give a brief sketch of what is to come. Surfaces will serve as a good example.

Suppose that **S** is a smooth surface. What we mean by that is that **S** has a nice tangent plane $\mathbf{T_p}$ at every point **p** that varies continuously as we move from point to point. Let us call the point where the tangent plane touches the surface its "origin." Since every tangent plane $\mathbf{T_p}$ is a two-dimensional vector space, we already know what it would mean to have an orientation $\sigma_\mathbf{p}$ for each $\mathbf{T_p}$ separately. The family of orientations $O = \{\sigma_\mathbf{p}\}$ is called an orientation for **S** if the orientations $\sigma_\mathbf{p}$ vary continuously from point to point. To explain what is meant by the notion of a continuously varying orientation, note that there is a well-defined one-to-one projection $\pi_\mathbf{p}$ of a neighborhood of the origin in $\mathbf{T_p}$ onto a neighborhood of **p** in the surface. Figure 1.14 shows

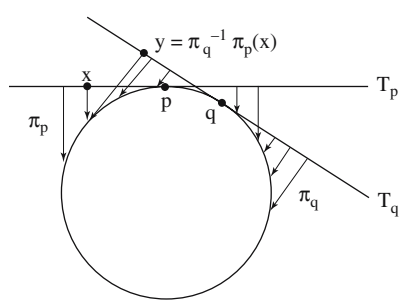



**Figure 1.14.** Defining continuously varying orientations.

this correspondence in the case of a curve. This means that if two points **p** and **q** are close, then the map

$$\pi_{\mathbf{p},\mathbf{q}} = \pi_{\mathbf{q}}^{-1}\pi_{\mathbf{p}}$$

is a well-defined bijection between a neighborhood of the origin in $\mathbf{T_p}$ and a neighborhood of the origin in $\mathbf{T_q}$. We can use this map to set up a correspondence between ordered bases in the two tangent spaces. In this way we can compare orientations, and we say that the orientations in O vary continuously if for nearby points $\sigma_{\mathbf{p}}$ and $\sigma_{\mathbf{q}}$ correspond under $\pi_{\mathbf{p},\mathbf{q}}$. An *oriented surface* is a pair (**S**,O), where **S** is a surface and O an orientation for **S**.

## 1.7 Convex Sets

**Definition.** A subset **X** of $\mathbf{R}^n$ is said to be *convex* if, for every pair of points **p** and **q** in **X**, the segment [**p**,**q**] is entirely contained in **X**.

Examples of convex and nonconvex sets are shown in Figure 1.15(a) and (b), respectively. The next proposition lists some basic facts about convex sets.

**1.7.1. Proposition.**

(1) Both the empty set and $\mathbf{R}^n$ are convex.
(2) Each halfplane in $\mathbf{R}^n$ is convex.
(3) The intersection of an arbitrary number of convex sets is convex.

**Proof.** Part (1) is trivial and parts (2) and (3) are left as exercises for the reader (Exercise 1.7.1 and 1.7.2).

Because convex sets have many nice properties, it is convenient to introduce the notion of the smallest convex set that contains a set.

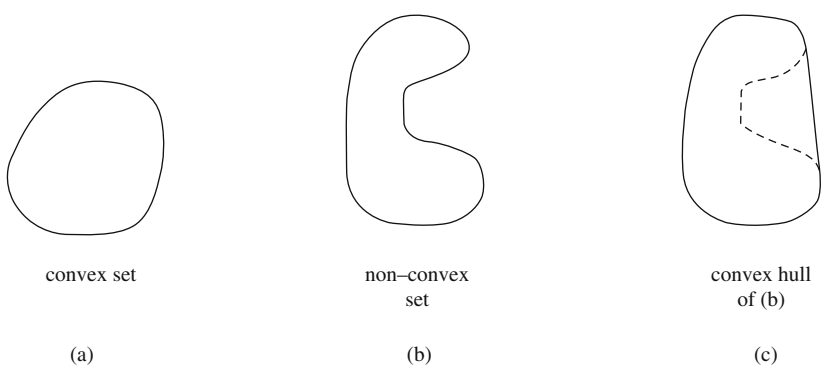



**Figure 1.15.** Convex and nonconvex sets and a convex hull.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$. The *convex hull* or *convex closure* of $\mathbf{X}$, denoted by conv($\mathbf{X}$), is defined by

$$\text{conv}(\mathbf{X}) = \cap \{\mathbf{C} \mid \mathbf{C} \text{ is a convex set that contains } \mathbf{X}\}.$$

This definition is similar to the one for the affine hull of a set. Two facts justify it. First, since each $\mathbf{R}^n$ is convex, we are never taking an empty intersection. Second, by Lemma 1.7.1(3) convex hulls are actually convex. One can also easily see that conv($\mathbf{X}$) is contained in any convex set that contains $\mathbf{X}$, which is why one refers to it as the "smallest" such set.

**1.7.2. Proposition.** If $\mathbf{X}$ is a convex subset of $\mathbf{R}^n$, then conv($\mathbf{X}$) = $\mathbf{X}$.

**Proof.** Exercise 1.7.3.

**Definition.** A bounded subset of $\mathbf{R}^n$ that is the intersection of a finite number of halfplanes is called a *convex linear polyhedron*.

The term "bounded" means that the set is contained in some closed disk about the origin. See Section 4.2. For example, we do not want to call $\mathbf{R}^n$ itself a convex linear polyhedron. A convex linear polyhedron is a special case of a linear polyhedron that will be defined in Section 6.3. It seems natural to give the definition here in order to show that the intersection of halfplanes produces many interesting and quite general sets and at the same time proves that these sets are convex. See Figure 1.16.

Certain convex linear polyhedra are especially interesting.

**Definition.** Let $k \geq 0$. A *k-dimensional simplex*, or *k-simplex*, is the convex hull $\boldsymbol{\sigma}$ of $k + 1$ linearly independent points $\mathbf{v}_0, \mathbf{v}_1, \ldots,$ and $\mathbf{v}_k$ in $\mathbf{R}^n$. We write $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k$. The points $\mathbf{v}_i$ are called the *vertices* of $\boldsymbol{\sigma}$. Often one writes $\boldsymbol{\sigma}^k$ to emphasize the dimension of $\boldsymbol{\sigma}$. If the dimension of $\boldsymbol{\sigma}$ is unimportant, then $\boldsymbol{\sigma}$ will be called simply a *simplex*. If $\{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_j\} \subseteq \{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k\}$, then $\boldsymbol{\tau} = \mathbf{w}_0\mathbf{w}_1 \cdots \mathbf{w}_j$ is called a *j-dimensional face* of $\boldsymbol{\sigma}$ and we shall write $\boldsymbol{\tau} \prec \boldsymbol{\sigma}$.

Figure 1.17 shows some examples of simplices and shows that our use of the term "k-dimensional" is justified. Note that $\mathbf{R}^2$ does not contain any three-dimensional

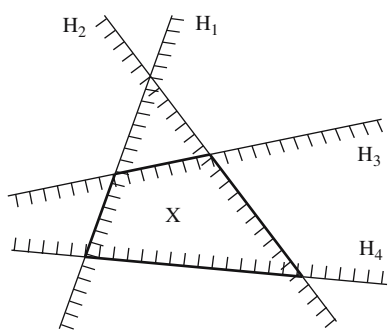



**Figure 1.16.** A convex linear polyhedron **X**.

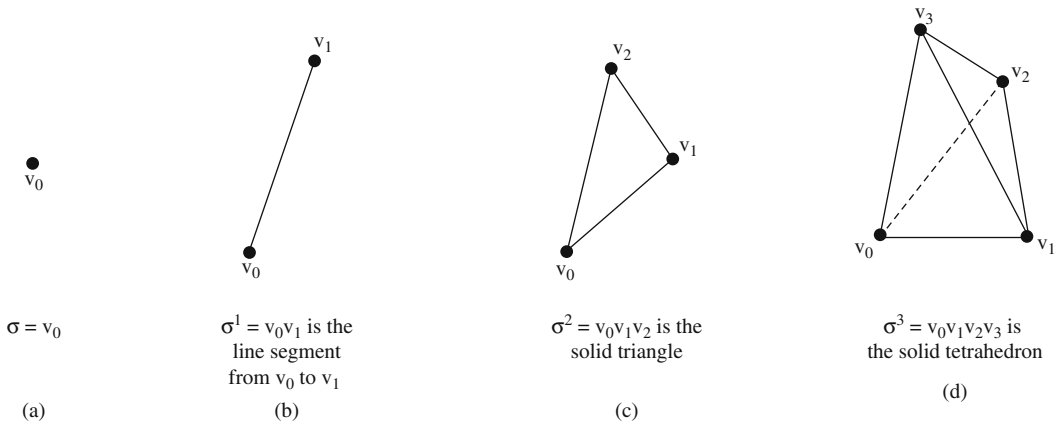



**Figure 1.17.** Some simplices.

simplex. In general, $\mathbf{R}^n$ contains at most n-dimensional simplices because it is not possible to find j linearly independent points in $\mathbf{R}^n$ for $j > n + 1$. Also, a simplex depends only on the set of vertices and not on their ordering. For example, $\mathbf{v}_0\mathbf{v}_1 = \mathbf{v}_1\mathbf{v}_0$. K-simplices are the simplest kind of building blocks for linear spaces called simplicial complexes, which are defined in Chapter 6, and they play an important role in algebraic topology. They have technical advantages over other regularly shaped regions such as cubes. In particular, their points have a nice representation as we shall show shortly in Theorem 1.7.4.

**1.7.3. Lemma.**

(1) The set $\mathrm{aff}(\{\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_k\})$ consists of the points $\mathbf{w}$ that can be written in the form

$$\mathbf{w} = \sum_{i=0}^{k} a_i\mathbf{v}_i, \quad \text{where} \quad \sum_{i=0}^{k} a_i = 1. \tag{1.25}$$

(2) The set $\mathrm{conv}(\{\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_k\})$ consists of the points $\mathbf{w}$ that can be written in the form

$$\mathbf{w} = \sum_{i=0}^{k} a_i\mathbf{v}_i, \quad \text{where} \quad a_i \in [0,1] \quad \text{and} \quad \sum_{i=0}^{k} a_i = 1.$$

**Proof.** To prove (1), let

$$\mathbf{S} = \left\{ \sum_{i=0}^{k} a_i\mathbf{v}_i \;\middle|\; \sum_{i=0}^{k} a_i = 1 \right\}.$$

If $\mathbf{w}$ belongs to aff($\{\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_k\}$), then we know from Theorem 1.4.4 that

$$\mathbf{w} = \mathbf{v}_0 + t_1\mathbf{v}_0\mathbf{v}_1 + \ldots + t_k\mathbf{v}_0\mathbf{v}_k \quad \text{for some } t_i \in \mathbf{R}.$$

This equation can be rewritten in the form

$$\mathbf{w} = (1 - t_1 - \ldots - t_k)\mathbf{v}_0 + t_1\mathbf{v}_1 + \ldots + t_k\mathbf{v}_k,$$

which shows that $\mathbf{w}$ belongs to $\mathbf{S}$. Conversely, if $\mathbf{w}$ belongs to $\mathbf{S}$, then

$$\mathbf{w} = \sum_{i=0}^{k} a_i\mathbf{v}_i \quad \text{for some } a_i \text{ such that} \quad \sum_{i=0}^{k} a_i = 1.$$

This equation can be rewritten in the form

$$\mathbf{w} = \mathbf{v}_0 + a_1\mathbf{v}_0\mathbf{v}_1 + \ldots + a_k\mathbf{v}_0\mathbf{v}_k.$$

Part (1) is proved.

To prove (2), let

$$\mathbf{S} = \left\{ \sum_{i=0}^{k} a_i\mathbf{v}_i \;\middle|\; a_i \in [0,1] \text{ and } \sum_{i=0}^{k} a_i = 1 \right\}.$$

We need to show that $\mathbf{S}$ is the smallest convex set containing $\{\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_k\}$. We show that $\mathbf{S}$ is convex first. Consider two points

$$\mathbf{w} = \sum_{i=0}^{k} a_i\mathbf{v}_i \quad \text{and} \quad \mathbf{w}' = \sum_{i=0}^{k} b_i\mathbf{v}_i$$

in $\mathbf{S}$ and let $t \in [0,1]$. Then

$$\begin{aligned}\mathbf{p} = t\mathbf{w} + (1-t)\mathbf{w}' &= t\left(\sum_{i=0}^{k} a_i\mathbf{v}_i\right) + (1-t)\left(\sum_{i=0}^{k} b_i\mathbf{v}_i\right) \\ &= \sum_{i=0}^{k} (t\,a_i + (1-t)b_i)\mathbf{v}_i.\end{aligned}$$

Clearly, $0 \le ta_i + (1 - t)b_i$. Furthermore,

$$\begin{aligned}\sum_{i=0}^{k} (t\,a_i + (1-t)b_i) &= t\left(\sum_{i=0}^{k} a_i\right) + (1-t)\left(\sum_{i=0}^{k} b_i\right) \\ &= t \cdot 1 + (1-t) \cdot 1 \\ &= 1.\end{aligned}$$

This also shows that $ta_i + (1 - t)b_i \le 1$; hence the point $\mathbf{p}$ belongs to $\mathbf{S}$, proving that $\mathbf{S}$ is convex since $\mathbf{p}$ is a typical point on the segment from $\mathbf{w}$ to $\mathbf{w}'$.

Next, we show that $\mathbf{S}$ belongs to every convex set $\mathbf{C}$ containing the points $\mathbf{v}_0$, $\mathbf{v}_1$, . . . , and $\mathbf{v}_k$. The case $k = 0$ is trivial. Assume that $k \ge 1$ and that the statement has been proved for all values smaller than k. Let

$$\mathbf{w} = \sum_{i=0}^{k} a_i \mathbf{v}_i$$

belong to $\mathbf{S}$. Since not all $a_i$ can be zero, we may assume without loss of generality that $a_0 \ne 0$. The case $a_0 = 1$ is trivial, and so assume that $a_0 < 1$. Thus we can write

$$\mathbf{w} = a_0\mathbf{v}_0 + (1 - a_0)\left(\sum_{i=1}^{k} \frac{a_i}{1 - a_0}\mathbf{v}_i\right).$$

But

$$\begin{aligned}\sum_{i=1}^{k} \frac{a_i}{1 - a_0} &= \frac{1}{1 - a_0}\sum_{i=1}^{k} a_i \\ &= \frac{1}{1 - a_0}(1 - a_0) \\ &= 1\end{aligned}$$

and $0 \le a_i/(1 - a_0) \le 1$. By our inductive hypothesis

$$\mathbf{u} = \sum_{i=1}^{k} \frac{a_i}{1 - a_0}\mathbf{v}_i$$

belongs to every convex set containing $\mathbf{v}_1$, $\mathbf{v}_2$, . . . , and $\mathbf{v}_k$. In particular, $\mathbf{u}$ belongs to $\mathbf{C}$. Since $\mathbf{v}_0$ belongs to $\mathbf{C}$, it follows that $\mathbf{w} = a_0\mathbf{v}_0 + (1 - a_0)\mathbf{u}$ belongs to $\mathbf{C}$ and we are done. Therefore,

$$\mathbf{S} = \operatorname{conv}(\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k\})$$

and (2) is proved.

An interesting consequence of Lemma 1.7.3(1) is that it gives us a homogeneous way of defining a plane. We could define a k-dimensional plane as a set defined by k + 1 linearly independent points $\mathbf{v}_0$, $\mathbf{v}_1$, . . . , $\mathbf{v}_k$ which satisfy equation (1.25) instead of the definition we gave in Section 1.5 that involved a point and a basis.

Lemma 1.7.3(2) motivates the following definition.

**Definition.** An expression of the form

$$\sum_{i=0}^{k} a_i \mathbf{v}_i, \quad \text{where} \quad a_i \in [0,1] \quad \text{and} \quad \sum_{i=0}^{k} a_i = 1,$$

and where the $\mathbf{v}_i$ are **any** objects for which the expression makes sense is called a *convex combination* of the $\mathbf{v}_i$.

**1.7.4. Theorem.** Let $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k$ be k + 1 linearly independent points.

(1) Every point $\mathbf{w}$ of $\operatorname{aff}(\{\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_k\})$ can be written **uniquely** in the form

$$\mathbf{w} = \sum_{i=0}^{k} a_i \mathbf{v}_i, \quad \text{where} \quad \sum_{i=0}^{k} a_i = 1.$$

(2) Every point $\mathbf{w}$ of the simplex $\sigma = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k$ can be written **uniquely** in the form

$$\mathbf{w} = \sum_{i=0}^{k} a_i \mathbf{v}_i, \quad \text{where} \quad a_i \in [0,1], \quad \text{and} \quad \sum_{i=0}^{k} a_i = 1.$$

Furthermore, the dimension and the vertices of a simplex are uniquely determined, that is, if $\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k = \mathbf{v}_0'\mathbf{v}_1' \cdots \mathbf{v}_t'$, then k = t and $\mathbf{v}_i = \mathbf{v}_i'$ after a renumbering of the $\mathbf{v}_i'$.

**Proof.** Lemma 1.7.3 showed that every point $\mathbf{w}$ has a representation as shown in (1) and (2). We need to show that it is unique. Suppose that we have two representations of the form

$$\mathbf{w} = \sum_{i=0}^{k} a_i \mathbf{v}_i = \sum_{i=0}^{k} a_i' \mathbf{v}_i.$$

Then

$$\begin{aligned}
\mathbf{0} = \mathbf{w} - \mathbf{w} &= \sum_{i=0}^{k} a_i \mathbf{v}_i - \sum_{i=0}^{k} a_i' \mathbf{v}_i \\
&= \sum_{i=0}^{k} \left(a_i - a_i'\right)\mathbf{v}_i \\
&= \sum_{i=0}^{k} \left(a_i - a_i'\right)(\mathbf{v}_i - \mathbf{v}_0) + \left(\sum_{i=0}^{k} \left(a_i - a_i'\right)\right)\mathbf{v}_0 \\
&= \sum_{i=0}^{k} \left(a_i - a_i'\right)(\mathbf{v}_i - \mathbf{v}_0) \\
&= \sum_{i=1}^{k} \left(a_i - a_i'\right)(\mathbf{v}_i - \mathbf{v}_0).
\end{aligned}$$

The second to last equality sign follows from the fact that

$$\sum_{i=0}^{k}\left(a_i - a_i'\right) = \sum_{i=0}^{k} a_i - \sum_{i=0}^{k} a_i' = 1 - 1 = 0.$$

But the vectors $\mathbf{v}_1 - \mathbf{v}_0$, $\mathbf{v}_2 - \mathbf{v}_0, \ldots,$ and $\mathbf{v}_k - \mathbf{v}_0$ are linearly independent, so that $a_i = a_i'$ for $i = 1,2, \ldots ,k$, which then also implies that $a_0 = a_0'$. This proves that the representation for $\mathbf{w}$ is unique.

The rest of part (2) is left as an exercise.

**Definition.** Using the notation in Theorem 1.7.4(1), the $a_i$ are called the *barycentric coordinates* of $\mathbf{w}$ with respect to the points $\mathbf{v}_i$. The point

$$\frac{1}{k+1}(\mathbf{v}_0 + \mathbf{v}_1 + \ldots + \mathbf{v}_k)$$

is called the *barycenter* of the simplex $\boldsymbol{\sigma}$.

**1.7.5. Example.** Let $\mathbf{v}_0 = (1,0)$, $\mathbf{v}_1 = (4,0)$, and $\mathbf{v}_2 = (3,5)$. We want to find the barycentric coordinates $(a_0,a_1,a_2)$ of $\mathbf{w} = (3,1)$ with respect to these vertices.

**Solution.** We must solve

$$a_0(1,0) + a_1(4,0) + a_2(3,5) = (3,1)$$

for $a_0$, $a_1$, and $a_2$. Since $a_2 = 1 - a_0 - a_1$, we really have to solve only two equations in two unknowns. The unique solutions are $a_0 = 4/15$, $a_1 = 8/15$, and $a_2 = 1/5$. The barycenter of the simplex $\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2$ is the point $(8/3,5/3)$.

Theorem 1.7.4 shows that barycentric coordinates are another way to parameterize points, which is why that terminology is used. They are a kind of weighted sum and are very useful in problems that deal with convex sets. In barycentric coordinates, the point $\mathbf{w}$ in the definition would be represented by the tuple $(a_0,a_1, \ldots ,a_k)$. The barycenter would have the representation

$$\left(\frac{1}{k+1}, \frac{1}{k+1}, \ldots, \frac{1}{k+1}\right).$$

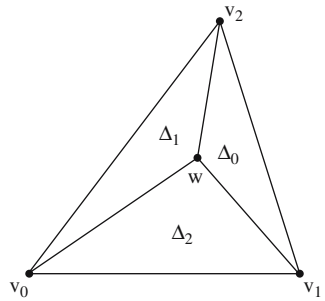
Barycentric coordinates give information about ratios of volumes (or areas in dimension 2). (For a general definition of volume in higher dimensions see Chapter 4.) Consider a simplex $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k$ and a point $\mathbf{w}$ in it. Let $(a_0,a_1, \ldots ,a_k)$ be the barycentric coordinates of $\mathbf{w}$. Let $\Delta$ be the volume of $\boldsymbol{\sigma}$ and let $\Delta_i$ be the volume of the simplex with vertices $\mathbf{v}_0, \mathbf{v}_1, \ldots , \mathbf{v}_{i-1}, \mathbf{w}, \mathbf{v}_{i+1}, \ldots , \mathbf{v}_k$. See Figure 1.18.

**1.7.6. Proposition.** $a_i = \dfrac{\Delta_i}{\Delta}$.

**Proof.** See [BoeP94].

Finally, barycentric coordinates are useful in describing linear maps between simplices. Let f be a map from the set of vertices of a simplex $\boldsymbol{\sigma}$ onto the set of vertices

**Figure 1.18.** Barycentric coordinates and volume ratios.




of another simplex $\boldsymbol{\tau}$. Let $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k$ and $\boldsymbol{\tau} = \mathbf{w}_0\mathbf{w}_1 \cdots \mathbf{w}_s$. If we express points of $\boldsymbol{\sigma}$ in terms of the (unique) barycentric coordinates with respect to its vertices, then f induces a well-defined map

$$|f| : \boldsymbol{\sigma} \to \boldsymbol{\tau}$$

defined by

$$|f|\left(\sum_{i=0}^{k} a_i \mathbf{v}_i\right) = \sum_{i=0}^{k} a_i f(\mathbf{v}_i).$$

**Definition.** The map $|f|$ is called the map from $\boldsymbol{\sigma}$ to $\boldsymbol{\tau}$ *induced* by the vertex map f.

In Chapter 6 we shall see that the map f is a special case of what is called a *simplicial map* between simplicial complexes and $|f|$ is the induced map on their underlying spaces. The main point to note here is that a map f of vertices induces a map $|f|$ on the whole simplex. (This is very similar to the way a map of basis vectors in a vector space induces a well-defined linear transformation of the whole vector space.) This gives us a simple abstract way to define linear maps between simplices, although a formula for this map in Cartesian coordinates is not that simple. See Exercises 1.7.6 and 1.7.7.

## 1.8 Principal Axes Theorems

The goal of this section is to state conditions under which a linear transformation can be diagonalized. We shall be dealing with vector spaces over **either** the reals **or** the complex numbers. We refer to the main theorems of this section as "principal axes theorems" because they can be interpreted as asserting the existence of certain coordinate systems (coordinate axes) with respect to which the transformation has a particularly simple description. Such diagonalization theorems are special cases of what are usually called "spectral theorems" in the literature because they deal with the eigenvalues (the "spectrum") of the transformation.

In the final analysis, it will turn out that a transformation is diagonalizable if the matrix associated with it is symmetric or Hermitian. Unfortunately, those properties of a matrix are not independent of the basis that is used to define the matrix. For example, it is possible to find a transformation and two bases, so that the matrix is symmetric with respect to one basis and not symmetric with respect to the other. The definition that captures the essence of the symmetry that we need is that of the "adjoint" transformation.

**1.8.1. Lemma.** Let **V** be an n-dimensional vector space over a field k. If $\alpha: \mathbf{V} \to k$ is a nonzero linear functional, then

$$\dim \ker(\alpha) = n - 1.$$

**Proof.** Since $\alpha$ is nonzero, $\dim \operatorname{im}(\alpha) = 1$ and so the lemma is an immediate consequence of Theorem B.10.3.

If the vector space **V** has an inner product $\bullet$, then it is easy to check that for each $\mathbf{u} \in \mathbf{V}$ the map

$$u^*: \mathbf{V} \to k$$

defined by

$$u^*(\mathbf{v}) = \mathbf{v} \bullet \mathbf{u}$$

is a linear transformation, that is, a linear functional. There is a converse.

**1.8.2. Theorem.** If $\alpha$ is a linear functional on an n-dimensional vector space **V** with inner product $\bullet$, then there is a unique **u** in **V**, so that

$$\alpha(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v}$$

for all **v** in **V**.

**Proof.** If $\alpha$ is the zero map, then **u** is clearly the zero vector. Assume that $\alpha$ is nonzero. Then by Lemma 1.8.1, the subspace $\mathbf{X} = \ker(\alpha)$ has dimension $n - 1$. Let $\mathbf{u}_0$ be any unit vector in the one-dimensional orthogonal complement $\mathbf{X}^\perp$ of **X**. We show that

$$\mathbf{u} = \overline{\alpha(\mathbf{u}_0)}\, \mathbf{u}_0$$

is the vector we are looking for. (The complex conjugate operation is needed in case we are dealing with vector spaces over the complex numbers.) If **v** is an arbitrary vector in **V**, then $\mathbf{V} = \mathbf{X} \oplus \mathbf{X}^\perp$ implies that $\mathbf{v} = \mathbf{x} + c\mathbf{u}$, for some **x** in **X** and some scalar c. But

$$\alpha(\mathbf{v}) = \alpha(\mathbf{x} + c\mathbf{u}) = \alpha(c\ \mathbf{u}) = c\ |\alpha(\mathbf{u}_0)|^2$$

and

$$\mathbf{u} \bullet \mathbf{v} = \mathbf{u} \bullet (\mathbf{x} + c\mathbf{u}) = c\mathbf{u} \bullet \mathbf{u} = c\ |\alpha(\mathbf{u}_0)|^2.$$

The existence part of the theorem is proved. To prove uniqueness, assume that there is another vector $\mathbf{u}'$ in $\mathbf{V}$ with $\alpha(\mathbf{v}) = \mathbf{u}' \bullet \mathbf{v}$. Then $(\mathbf{u} - \mathbf{u}') \bullet \mathbf{v} = 0$ for all $\mathbf{v}$ in $\mathbf{V}$. In particular, letting $\mathbf{v} = \mathbf{u} - \mathbf{u}'$, we get that

$$(\mathbf{u} - \mathbf{u}') \bullet (\mathbf{u} - \mathbf{u}') = 0,$$

which implies that $\mathbf{u} = \mathbf{u}'$ and we are done.

Next, assume that $\mathbf{V}$ is a vector space and $T : \mathbf{V} \to \mathbf{V}$ is a linear transformation. Given $\mathbf{v} \in \mathbf{V}$, define a linear functional $T_{\mathbf{v}}$ by

$$T_{\mathbf{v}}(\mathbf{w}) = T(\mathbf{w}) \bullet \mathbf{v}.$$

By Theorem 1.8.2, there is a unique vector $\mathbf{v}^*$, so that

$$T_{\mathbf{v}}(\mathbf{w}) = \mathbf{v}^* \bullet \mathbf{w}.$$

**Definition.** The map

$$T^* : \mathbf{V} \to \mathbf{V}$$

defined by

$$T^*(\mathbf{v}) = \mathbf{v}^*$$

is called the *adjoint* of T.

**1.8.3. Lemma.** The adjoint map $T^*$ satisfies

$$T(\mathbf{v}) \bullet \mathbf{w} = \mathbf{v} \bullet T^*(\mathbf{w})$$

for all $\mathbf{v}, \mathbf{w} \in \mathbf{V}$.

**Proof.** By definition, $T(\mathbf{v}) \bullet \mathbf{w} = T_{\mathbf{w}}(\mathbf{v}) = \mathbf{w}^* \bullet \mathbf{v} = T^*(\mathbf{w}) \bullet \mathbf{v}$.

**1.8.4. Lemma.** The adjoint map $T^*$ is a linear transformation.

**Proof.** Using Lemma 1.8.3 and the linearity of the dot product, we have that

$$\begin{aligned}
\mathbf{u} \bullet T^*(a\ \mathbf{v} + b\ \mathbf{w}) &= T(\mathbf{u}) \bullet (a\ \mathbf{v} + b\ \mathbf{w}) \\
&= a\ T(\mathbf{u}) \bullet \mathbf{v} + b\ T(\mathbf{u}) \bullet \mathbf{w} \\
&= a\ \mathbf{u} \bullet T^*(\mathbf{v}) + b\ \mathbf{u} \bullet T^*(\mathbf{w}) \\
&= \mathbf{u} \bullet (a\ T^*(\mathbf{v}) + b\ T^*(\mathbf{w})).
\end{aligned}$$

Since this holds for all vectors $\mathbf{u}$, we must have $T^*(a\mathbf{v} + b\mathbf{w}) = aT^*(\mathbf{v}) + bT^*(\mathbf{w})$, which proves the lemma.

**Definition.** A linear transformation $T: \mathbf{V} \to \mathbf{V}$ is called *self-adjoint* if $T = T^*$.

It follows from Lemma 1.8.3 that if a linear transformation T is self-adjoint then

$$T(\mathbf{v}) \bullet \mathbf{w} = \mathbf{v} \bullet T(\mathbf{w}).$$

One can prove the converse, namely

**1.8.5. Lemma.** If a linear transformation $T: \mathbf{V} \to \mathbf{V}$ satisfies

$$T(\mathbf{v}) \bullet \mathbf{w} = \mathbf{v} \bullet T(\mathbf{w}).$$

for all $\mathbf{v}, \mathbf{w} \in \mathbf{V}$, then T is self-adjoint.

**Proof.** The lemma follows from the fact that for all vectors $\mathbf{v}$ we have

$$\mathbf{v} \bullet T(\mathbf{w}) = T(\mathbf{v}) \bullet \mathbf{w} = \mathbf{v} \bullet T^*(\mathbf{w}).$$

**1.8.6. Theorem.** Let $\mathbf{V}$ be a real vector space and $T: \mathbf{V} \to \mathbf{V}$ a linear transformation. If M is the matrix of T with respect to an orthonormal basis, then the matrix of the adjoint $T^*$ of T with respect to that same basis is $M^T$.

**Proof.** Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be an orthonormal basis. By definition of the matrix for a linear transformation and properties of orthonormal bases, the ijth entries of the matrices for T and $T^*$ are $T(\mathbf{u}_i) \cdot \mathbf{u}_j$ and $T^*(\mathbf{u}_i) \cdot \mathbf{u}_j$, respectively. But

$$T(\mathbf{u}_i) \bullet \mathbf{u}_j = \mathbf{u}_i \bullet T^*(\mathbf{u}_j) = T^*(\mathbf{u}_j) \bullet \mathbf{u}_i.$$

**1.8.7. Corollary.** The matrix for a self-adjoint linear transformation on a real vector space with respect to an orthonormal basis is symmetric. Conversely, if the matrix for a linear transformation over a real vector space with respect to an orthonormal basis is symmetric, then the linear transformation is self-adjoint.

**Proof.** This is an easy consequence of Theorem 1.8.6.

Self-adjoint transformations are sometimes called *symmetric* transformations because of Corollary 1.8.7. The complex analogs of Theorem 1.8.6 and Corollary 1.8.7 simply replace the transpose with the complex conjugate transpose and the self-adjoint transformations in this case are sometimes called *Hermitian*.

We now return to the problem of when a linear transformation can be diagonalized. We shall deal with real and complex vector spaces separately. The reason is that eigenvalues are roots of the characteristic polynomial of a transformation. Although polynomials always factor completely into linear factors over the complex numbers,

this is not always the case over the reals. A polynomial may have no roots at all over the reals.

**1.8.8. Lemma.** Every eigenvalue of a self-adjoint linear transformation T on a complex vector space **V** with inner product • is real.

**Proof.** Let $\lambda$ be an eigenvalue for T and **u** a nonzero eigenvector for $\lambda$. Then

$$\lambda(\mathbf{u} \bullet \mathbf{u}) = \lambda\mathbf{u} \bullet \mathbf{u} = T(\mathbf{u}) \bullet \mathbf{u} = \mathbf{u} \bullet T^*(\mathbf{u}) = \mathbf{u} \bullet T(\mathbf{u}) = \mathbf{u} \bullet \lambda\mathbf{u} = \overline{\lambda}(\mathbf{u} \bullet \mathbf{u}).$$

Since $\mathbf{u} \bullet \mathbf{u} \neq 0$, $\lambda = \overline{\lambda}$, that is, $\lambda$ is real.

**1.8.9. Lemma.** Let T be a self-adjoint linear transformation over a real n-dimensional vector space **V**, $n \geq 1$, with inner product •. Then

(1) The characteristic polynomial of T is a product of linear factors.
(2) Eigenvectors corresponding to distinct eigenvalues are orthogonal.

**Proof.** By passing to the matrix A for T, (1) follows immediately from Lemma 1.8.8 because we can think of A as defining a complex transformation on $\mathbf{C}^n$ and every polynomial of degree n factors into linear factors over the complex numbers. To prove (2), assume that $T(\mathbf{u}) = \lambda\mathbf{u}$ and $T(\mathbf{v}) = \mu\mathbf{v}$ for $\lambda \neq \mu$. Then

$$\lambda(\mathbf{u} \bullet \mathbf{v}) = \lambda\mathbf{u} \bullet \mathbf{v} = T(\mathbf{u}) \bullet \mathbf{v} = \mathbf{u} \bullet T(\mathbf{v}) = \mathbf{u} \bullet \mu\mathbf{v} = \mu(\mathbf{u} \bullet \mathbf{v}).$$

Since $\lambda \neq \mu$, it follows that $\mathbf{u} \bullet \mathbf{v} = 0$, and we are done.

**1.8.10. Theorem.** (The Real Principal Axes Theorem) Let T be a self-adjoint transformation on an n-dimensional real vector space **V**, $n \geq 1$. Then **V** admits an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ consisting of eigenvectors of T, that is,

$$T(\mathbf{u}_i) = \lambda_i \mathbf{u}_i$$

for some real numbers $\lambda_i$.

**Proof.** The proof is by induction on n. The theorem is clearly true for $n = 1$. Assume, therefore, inductively that it has been proved for dimension $n - 1$, $n > 1$. There are basically two steps involved in the rest of the proof.

First, we need to know that the transformation actually has at least one real eigenvalue $\lambda$. This was proved by Lemma 1.8.9(1). Let **v** be a nonzero eigenvector for $\lambda$ and let $\mathbf{u}_1 = \mathbf{v}/|\mathbf{v}|$.

The second step, in order to use the inductive hypothesis, is to show that the orthogonal complement $\mathbf{W}^\perp$ of $\mathbf{W} = \langle\mathbf{v}\rangle = \langle\mathbf{u}_1\rangle$ is an invariant subspace of T. This follows from the fact that if $\mathbf{w} \in \mathbf{W}^\perp$, then

$$\mathbf{v} \bullet T(\mathbf{w}) = v \bullet T^*(\mathbf{w}) = T(\mathbf{v}) \bullet \mathbf{w} = \lambda\mathbf{v} \bullet \mathbf{w} = 0,$$

so that $T(\mathbf{w}) \in \mathbf{W}^{\perp}$. Clearly, $S = T|\mathbf{W}^{\perp}$ is a self-adjoint transformation on the $(n-1)$-dimensional vector space $\mathbf{W}^{\perp}$. The inductive hypothesis applied to S means that there is an orthonormal basis $\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n$ for $\mathbf{W}^{\perp}$ which are eigenvectors for S (and hence for T). The vectors $\mathbf{u}_i$ are obviously what we wanted, proving the theorem.

In the case of Theorem 1.8.10, the name "Principal Axes Theorem" comes from its role in finding the principal axes of ellipses. The matrix form of Theorem 1.8.10 is

**1.8.11. Theorem.** If A is a real symmetric $n \times n$ matrix, then there exists an orthogonal matrix P so that $D = P^{-1}AP$ is a diagonal matrix. In particular, every real symmetric matrix is similar to a diagonal one.

**Proof.** Simply let the columns of P be the vectors that form an orthonormal basis of eigenvectors.

Note that Theorem 1.8.11 only gives sufficient conditions for a matrix to be similar to a diagonal one. Nonsymmetric matrices can also be similar to a diagonal one. For necessary and sufficient conditions for a matrix to be diagonalizable see Theorem C.4.10.

In Theorem 1.8.11, the number s of positive diagonal entries of D is uniquely determined by A. We may assume that the diagonal of D has the s positive entries first, followed by $r - s$ negative entries, followed by $n - r$ zeros, where r is the rank of A.

**1.8.12. Example.** Let

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

We want to find an orthogonal matrix P so that $P^{-1}AP$ is a diagonal matrix.

**Solution.** Consider A to be the matrix of a linear transformation T on $\mathbf{R}^2$. Now, the roots of the characteristic polynomial

$$\det(t\,I^2 - A) = t^2 - 4t + 3$$

are 1 and 3, which are the eigenvalues of T. To find the corresponding eigenvectors, we must solve

$$(x\ y)(I^2 - A) = \mathbf{0}$$

and

$$(x\ y)(3\,I^2 - A) = \mathbf{0}.$$

This leads to two pairs of equations

$$\begin{aligned} -x + y &= 0 \\ x - y &= 0 \end{aligned}$$

and

$$\begin{aligned} x+y&=0\\ x+y&=0. \end{aligned}$$

In other words, the vectors $\mathbf{v}_1 = (1,1)$ and $\mathbf{v}_2 = (1,-1)$ are eigenvectors corresponding to eigenvalues 1 and 3, respectively. Let $\mathbf{u}_1 = (1/\sqrt{2},1/\sqrt{2})$ and $\mathbf{u}_2 = (1/\sqrt{2},-1/\sqrt{2})$. If P is the matrix with columns $\mathbf{u}_i$, then

$$P=\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{and} \quad P^{-1}AP=\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

**1.8.13. Example.** Let

$$A=\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

We want to find an orthogonal matrix P so that $P^{-1}AP$ is a diagonal matrix.

**Solution.** The roots of the characteristic polynomial

$$\det(t\,I^3 - A)=\det\begin{pmatrix} t-2 & -1 & -1 \\ -1 & t-2 & -1 \\ -1 & -1 & t-2 \end{pmatrix}=(t-1)^2(t-4)$$

are 1 and 4. To find the eigenvectors corresponding to the eigenvalue 1 we need to solve the equations

$$\begin{aligned} -x-y-z&=0\\ -x-y-z&=0\\ -x-y-z&=0. \end{aligned}$$

The solution set $\mathbf{X}$ has the form

$$\{(-y-z,y,z) \mid y,z \in \mathbf{R}\} = \{y(-1,1,0)+z(-1,0,1) \mid y,z \in \mathbf{R}\}.$$

Applying the Gram-Schmidt algorithm to the basis $(-1,1,0)$, $(-1,0,1)$ produces the orthonormal basis $\mathbf{u}_1 = (-1/\sqrt{2},1/\sqrt{2},0)$ and $\mathbf{u}_2 = (-1/\sqrt{6},-1/\sqrt{6},2/\sqrt{6})$ for $\mathbf{X}$. Next, to find the eigenvector for the eigenvalue 4, we need to solve

$$\begin{aligned} 2x-y-z&=0\\ -x+2y-z&=0\\ -x-y+2z&=0. \end{aligned}$$

The solutions to these equations have the form x(1,1,1). Let $\mathbf{u}_3 = (1/\sqrt{3},1/\sqrt{3},1/\sqrt{3})$. Finally, if P is the matrix whose columns are the $\mathbf{u}_i$, then

$$P = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \quad \text{and} \quad P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

**Definition.** A linear transformation T is said to be *normal* if it commutes with its adjoint, that is, $TT^* = T^*T$.

**1.8.14. Theorem.** (The Complex Principal Axes Theorem) Let T be a normal transformation on an n-dimensional complex vector space **V**, $n \geq 1$. Then **V** admits an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ consisting of eigenvectors of T, that is,

$$T(\mathbf{u}_i) = \lambda_i \mathbf{u}_i$$

for some complex numbers $\lambda_i$.

**Proof.** See [Lips68].

The matrix form of Theorem 1.8.14 is

**1.8.15. Theorem.** If A is a normal matrix, then there exists an unitary matrix P so that $D = P^{-1}AP$ is a diagonal matrix.

**Proof.** See [Lips68].

## 1.9 Bilinear and Quadratic Maps

This section describes some maps that appear quite often in mathematics. However, we are not interested in just the general theory. Quadratic maps and quadratic forms, in particular, have important applications in a number of areas of geometry and topology. For example, the conics, which are an important class of spaces in geometry, are intimately connected with quadratic forms. Other applications are found in Chapters 8 and 9.

**Definition.** A *bilinear map* on a vector space **V** over a field k is a function $f: \mathbf{V} \times \mathbf{V} \to k$ satisfying

(1) $f(a\mathbf{v} + b\mathbf{v}',\mathbf{w}) = af(\mathbf{v},\mathbf{w}) + bf(\mathbf{v}',\mathbf{w})$, and
(2) $f(\mathbf{v},a\mathbf{w} + b\mathbf{w}') = af(\mathbf{v},\mathbf{w}) + bf(\mathbf{v},\mathbf{w}')$,

for all $\mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in \mathbf{V}$ and $a, b \in k$.

**1.9.1. Example.** The dot product on $\mathbf{R}^n$ is a bilinear map. More generally, an inner product is a bilinear map.

**1.9.2. Example.** The determinant function on $\mathbf{R}^2$ (where we think of either the rows or columns of the matrix as vectors in $\mathbf{R}^2$) is a bilinear map.

**1.9.3. Example.** Let A be an $n \times n$ matrix. The map $f: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ defined by

$$f(\mathbf{v}, \mathbf{w}) = \mathbf{v}\, A\, \mathbf{w}^T$$

is a bilinear map.

**1.9.4. Example.** Let T be a linear transformation on $\mathbf{R}^n$. The map $f: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ defined by

$$f(\mathbf{v}, \mathbf{w}) = \mathbf{v} \bullet T(\mathbf{w})$$

is a bilinear map.

**Definition.** Let f be a bilinear map on a vector space $\mathbf{V}$. Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis for $\mathbf{V}$ and let $a_{ij} = f(\mathbf{v}_i, \mathbf{v}_j)$. The matrix $A = (a_{ij})$ is called the *matrix for f* with respect to the basis B. The determinant of A is called the *discriminant* of f with respect to the basis B.

The matrix for a bilinear map clearly depends on the chosen basis. However, the following is true:

**1.9.5. Proposition.** If $B' = (\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_n')$ is another ordered basis for $\mathbf{V}$ and if A′ is the matrix of the bilinear map f with respect to B′, then

$$A' = CAC^T,$$

where $C = (c_{ij})$ is the matrix relating the basis B to the basis B′, that is,

$$\mathbf{v}_i' = \sum_{j=1}^{n} c_{ij} \mathbf{v}_j.$$

**Proof.** This can be checked by a straightforward computation.

**Definition.** A real $n \times n$ matrix A is said to be *congruent* to a real $n \times n$ matrix B if there exists a nonsingular matrix C such that $A = CBC^T$.

It is easy to show that the congruence relation is an equivalence relation on the set of all $n \times n$ real matrices. We can rephrase Proposition 1.9.5.

**1.9.6. Corollary.** The matrix of a bilinear map is unique up to congruence, so that the study of bilinear maps is equivalent to the study of congruence classes of matrices.

**Definition.** The *rank* of a bilinear map f is the rank of any matrix for f. A bilinear map on an n-dimensional vector space is said to be *degenerate* or *nondegenerate* if its rank is less than n or equal to n, respectively.

It follows from Proposition 1.9.5 that the rank of a bilinear map is well defined.

**1.9.7. Proposition.** The matrix associated to a symmetric bilinear map with respect to any basis is a symmetric matrix.

**Proof.** Exercise.

**Definition.** A *quadratic map* on a vector space $\mathbf{V}$ over a field k is any map $q:\mathbf{V} \to k$ that can be defined in the form $q(\mathbf{v}) = f(\mathbf{v},\mathbf{v})$, where f is some bilinear map on $\mathbf{V}$. In that case, q is also called the *quadratic map associated to f*. The quadratic map q is said to be *degenerate* or *nondegenerate* if f is. A *discriminant* of q is defined to be discriminant of f with respect to some basis for $\mathbf{V}$. The quadratic map q and the bilinear map f are said to be *positive definite* if $q(\mathbf{v}) = f(\mathbf{v},\mathbf{v}) > 0$ for all $\mathbf{v} \neq \mathbf{0}$.

**1.9.8. Example.** Let f be a bilinear map on $\mathbf{R}^2$ and assume that

$$f(\mathbf{v},\mathbf{w}) = a_{11}v_1w_1 + a_{12}v_1w_2 + a_{21}v_2w_1 + a_{22}v_2w_2.$$

The quadratic map q associated to f is then given by

$$q(\mathbf{v}) = a_{11}v_1^2 + (a_{12} + a_{21})v_1v_2 + a_{22}v_2^2.$$

We see that q is just a homogeneous polynomial of degree 2 in $v_1$ and $v_2$.

If the field k does not have characteristic 2 and if f is a symmetric bilinear map with associated quadratic map q, then

$$f(\mathbf{v},\mathbf{w}) = \frac{1}{2}[q(\mathbf{v}+\mathbf{w}) - q(\mathbf{v}) - q(\mathbf{w})].$$

In other words, knowledge of q alone allows one to reconstruct f, so that the concepts "symmetric bilinear map" and "quadratic map" are really just two ways of looking at the same thing.

The form of the quadratic map in Example 1.9.8 and others like it motivates what is basically nothing but some alternate terminology for talking about quadratic maps.

**Definition.** A *d-ic form* over a field k is a homogeneous polynomial over k of degree d in an appropriate number of variables. A *linear* or *quadratic form* is a d-ic form where d is 1 or 2, respectively.

For example, 2x + 3y is a linear form in variables x and y and

$$x^2 + 5y^2 - 2z^2 + 3xy + yz$$

is a quadratic form in variables x, y, and z. Note that the quadratic form can be rewritten with symmetric cross-terms as

$$x^2 + 5y^2 - 2z^2 + \frac{3}{2}xy + \frac{3}{2}yx + \frac{1}{2}yz + \frac{1}{2}zy.$$

It follows that one can associate the symmetric matrix

$$\begin{pmatrix} 1 & \frac{3}{2} & 0 \\ \frac{3}{2} & 5 & \frac{1}{2} \\ 0 & \frac{1}{2} & -2 \end{pmatrix}$$

with this form. More generally, if the field k does not have characteristic 2, such as, for example, **R** or **C**, we can make the cross-terms symmetric with the trick shown in the example above. It follows that in this case **every** quadratic form in n variables is simply an expression of the type

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j,$$

where $A = (a_{ij})$ is a symmetric matrix. This means that every quadratic form defines a unique quadratic map

$$q: \mathbf{V} \to k$$

on a vector space **V** in the following way: Choose an ordered basis $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ for **V**. Let $\mathbf{v} \in \mathbf{V}$ and suppose that

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i.$$

Then

$$q(\mathbf{v}) = \mathbf{x}\, A\, \mathbf{x}^T = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j,$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. In this case also, the matrix for the associated bilinear map is just the matrix A. Of course, for all this to make sense we are treating the $x_i$ as values rather than variables, but we can see that from a theoretical point of view there is no difference between the theory of quadratic forms and quadratic maps. This explains why in the literature the terms "quadratic map" and "quadratic form" are

often used interchangeably. In particular, one uses the same terms, such as "degenerate," "nondegenerate," "positive definite," or "discriminant" for both. One will sometimes also find the term "bilinear form" used instead of "bilinear map."

**Note.** In this book we shall often use the more popular term "quadratic form" even though we may interpret it as a quadratic map because that is typically more convenient computationally. Specifically, when the field k does not have characteristic 2, we shall always feel free to switch between a quadratic form and the appropriate corresponding quadratic map with its unique associated symmetric bilinear map whose symmetric matrix is unique up to congruence.

Now, an arbitrary quadratic form can be quite complicated. Key to understanding them is the fact that one can always choose a basis, so that with respect to this basis, the form has a nice simple structure.

**1.9.10. Theorem.** (The Principal Axes Theorem) Given a quadratic form q defined on $\mathbf{R}^n$, there exists an orthonormal basis for $\mathbf{R}^n$ with respect to which q has the form

$$q(x_1, x_2, \ldots, x_n) = \lambda_1 x_1^2 + \ldots + \lambda_s x_s^2 - \lambda_{s+1} x_{s+1}^2 - \ldots - \lambda_{s+t} x_{s+t}^2, \quad \text{where} \quad \lambda_i > 0.$$

The difference s – t is called the *signature* of the quadratic form or the associated symmetric bilinear map.

**Proof.** This is an immediate consequence of Theorem 1.8.11. The integers s and t, and hence the signature, are independent of the basis and hence invariants of the quadratic form.

If we do not insist on an orthonormal basis for the diagonalization of a quadratic form, then there is a weaker version of Theorem 1.9.10. It is interesting because there is a simpler algorithm for finding a diagonalizing basis for a quadratic form. Here is its matrix form.

**1.9.11. Theorem.** If A is a real symmetric $n \times n$ matrix of rank r, then A is congruent to a **unique** diagonal matrix whose first s diagonal entries are +1, the next r – s entries are –1, and the remaining entries are zeros.

**Proof.** We sketch a proof. For more details, see [Fink72]. Assume that A is not the zero matrix; otherwise, there is nothing to prove.

**Step 1.** To make A congruent to a matrix $A_1$ that has a nonzero diagonal element.

If A has a nonzero diagonal element, then let $A_1 = A$. If all diagonal elements of A are zero, let $a_{ij}$ be any nonzero entry of A. Let E be the elementary matrix $E_{ji}(1)$, which has 1s on the diagonal, a 1 in the jith place, and zeros everywhere else. Let $A_1 = EAE^T$. The matrix $A_1$ is obtained from A by adding the jth row of A to the ith row followed by adding the jth column of the result to the ith column. It is easy to see that the ith diagonal element of $A_1$ is $2a_{ij}$ and hence nonzero.

**Step 2.** To make $A_1$ congruent to a matrix $A_2$, which has $a_{11}$ nonzero.

Let $F = (f_{ij})$ be the elementary matrix defined by

$$\begin{aligned} f_{st} &= 1, \quad \text{if} \quad s = t, \quad s \neq 1 \text{ or } i, \\ &= 1, \quad \text{if} \quad s = 1, \quad t = i \\ &= 1, \quad \text{if} \quad s = i, \quad t = 1 \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

Then $A_2 = FA_1F^T$ is the matrix obtained from $A_1$ by interchanging the first and ith diagonal element.

**Step 3.** To make $A_2$ congruent to a matrix $A_3$ in which the only nonzero element in the first row or first column is $a_{11}$.

Step 3 is accomplished via elementary matrices like in Step 1 that successively add multiples of the first row to all the other rows from 2 to n and the same multiples of the first column to the other columns.

After Step 3, the matrix $A_3$ will have the form

$$\begin{pmatrix} a_{11} & 0 \\ 0 & B \end{pmatrix}$$

where B is a symmetric $(n - 1) \times (n - 1)$ matrix. Repeating Steps 1–3 on the matrix B and so on will show that A is congruent to a diagonal matrix with the first r diagonal entries nonzero. By interchanging the diagonal entries like in Step 2 if necessary, we may assume that all the positive entries come first. This shows that A is congruent to a diagonal matrix

$$G = D(d_1, \ldots, d_s, -d_{s+1}, \ldots, -d_r, 0, \ldots, 0),$$

where $d_i > 0$. If

$$H = D\left(\frac{1}{\sqrt{d_1}}, \ldots, \frac{1}{\sqrt{d_r}}, 0, \ldots, 0\right),$$

then $HGH^T$ has the desired form. To see why s is uniquely determined see [Fink72].

One nice property of the proof of Theorem 1.9.11 is that it is constructive.

**1.9.12. Example.** To show that the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 0 \end{pmatrix}$$

is congruent to a diagonal one with ±1s or 0 on the diagonal.

**Solution.** We follow the steps outlined in Theorem 1.9.11. If the elementary matrices E, F, and G are defined by

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix},$$

then

$$A_1 = G\,F\,E\,A\,E^T\,F^T\,G^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Finally, define the elementary diagonal matrix H by

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and observe that

$$A_2 = H\,A_1\,H^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Therefore, if M = HGFE, then $MAM^T = A_2$ is the desired matrix and we are done.

It is worth pointing out one consequence of Theorem 1.9.10.

**1.9.13. Corollary.** A positive definite quadratic form is nondegenerate and all of its discriminants are positive.

**Proof.** We may assume that the vector space is $\mathbf{R}^n$ and then must have s = n in Theorem 1.9.10. The discriminant is certainly positive with the respect to the orthonormal basis guaranteed by the theorem. The reason that the discriminant is always positive is that the determinant of congruent matrices differs by a square.

## 1.10 The Cross Product Reexamined

In Section 1.5 we observed that $\mathbf{R}^3$ has not only a dot product but also a cross product. Note that the cross product produces another vector, whereas the dot product was a real number. Various identities involving the dot and cross product are known. The

cross product is a "product" that behaves very much like the product in the case of real numbers except that it is not commutative. The two operations of vector addition and the cross product make $\mathbf{R}^3$ into a (noncommutative) **ring**. Is there a similar product in other dimensions? Unfortunately not, but the cross product does arise from a general construction that applies to all dimensions and that is worth looking at because it will give us additional insight into the cross product.

**1.10.1. Theorem.** Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1} \in \mathbf{R}^n$. Define a map $T: \mathbf{R}^n \to \mathbf{R}$ by

$$T(\mathbf{w}) = \det\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{n-1} \\ \mathbf{w} \end{pmatrix}.$$

Then there is a unique $\mathbf{u} \in \mathbf{R}^n$ such that $T(\mathbf{w}) = \mathbf{u} \bullet \mathbf{w}$ for all $\mathbf{w}$.

**Proof.** This theorem is an immediate corollary to Theorem 1.8.2 because properties of the determinant function show that T is a linear functional.

**Definition.** Using the notation of Theorem 1.10.1, the vector $\mathbf{u}$ is called the (*generalized*) *cross product* of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{n-1}$ and is denoted by $\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1}$.

**1.10.2. Proposition.** The generalized cross product satisfies the following basic properties:

(1) It is commutative up to sign, that is,

$$\mathbf{v}_{\sigma(1)} \times \mathbf{v}_{\sigma(2)} \times \cdots \times \mathbf{v}_{\sigma(n-1)} = \operatorname{sign}(\sigma)\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1}$$

for all permutations $\sigma$ of $\{1, 2, \ldots, n-1\}$.

(2) It is a multi-linear map, that is,

$$\mathbf{v}_1 \times \cdots \times a\,\mathbf{v}_i \times \cdots \times \mathbf{v}_{n-1} = a(\mathbf{v}_1 \times \cdots \times \mathbf{v}_i \times \cdots \times \mathbf{v}_{n-1})$$
$$\mathbf{v}_1 \times \cdots \times (\mathbf{v}_i + \mathbf{v}_i') \times \cdots \times \mathbf{v}_{n-1} = (\mathbf{v}_1 \times \cdots \times \mathbf{v}_i \times \cdots \times \mathbf{v}_{n-1}) + (\mathbf{v}_1 \times \cdots \times \mathbf{v}_i' \times \cdots \times \mathbf{v}_{n-1})$$

(3) $(\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1}) \bullet \mathbf{v}_i = 0$ , for all i.

(4) If the vectors $\mathbf{v}_i$ are linearly independent, then the ordered basis

$$(\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_{n-1}, \mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1})$$

induces the standard orientation on $\mathbf{R}^n$.

**Proof.** Facts (1) and (2) are immediate from the definition using properties of the determinant. Fact (3) follows from the observation that the determinant of a matrix with two equal rows is zero, so that each $\mathbf{v}_i$ lies in the kernel of T in Theorem 1.10.1.

To prove (4), note that by definition

$$\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1} \bullet \mathbf{w} = \det \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_n \\ \mathbf{w} \end{pmatrix} \tag{1.26}$$

holds for **all** vectors $\mathbf{w}$. When $\mathbf{w}$ is the vector $\mathbf{v}_1 \times \mathbf{v}_2 \times \cdots \times \mathbf{v}_{n-1}$ we see that the left-hand side of (1.26) is positive, which implies that the determinant is also. Now use Lemma 1.6.4.

**1.10.3. Proposition.** In $\mathbf{R}^3$ the generalized cross product agrees with the usual cross product as defined in Section 1.5.

**Proof.** This is Exercise 1.10.1.

The next proposition lists a few of the well-known properties of the cross product in the special case of $\mathbf{R}^3$.

**1.10.4. Proposition.** The (generalized) cross product in $\mathbf{R}^3$ satisfies

(1) $|\mathbf{u} \times \mathbf{v}| = |\mathbf{v}|\ |\mathbf{w}| \sin\theta$, where $\theta$ is the angle between $\mathbf{u}$ and $\mathbf{v}$.
(2) $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \bullet \mathbf{w})\ \mathbf{v} - (\mathbf{u} \bullet \mathbf{v})\ \mathbf{w}$
$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \bullet \mathbf{w})\ \mathbf{v} - (\mathbf{v} \bullet \mathbf{w})\ \mathbf{u}$
(3) $|\mathbf{u} \times \mathbf{v}|^2 = |\mathbf{u}|^2\ |\mathbf{v}|^2 - (\mathbf{u} \bullet \mathbf{v})^2$
(4) $(\mathbf{u}_1 \times \mathbf{u}_2) \bullet (\mathbf{v}_1 \times \mathbf{v}_2) = (\mathbf{u}_1 \bullet \mathbf{v}_1)(\mathbf{u}_2 \bullet \mathbf{v}_2) - (\mathbf{u}_1 \bullet \mathbf{v}_2)(\mathbf{u}_2 \bullet \mathbf{v}_1)$

**Proof.** Exercise 1.10.2.

One way to look at identity (3) in Proposition 1.10.4 is that the cross product measures the deviation from equality in the Cauchy-Schwarz inequality.

**1.10.5. Example.** Find the equation of the plane through (1,0,3) with basis $\mathbf{v}_1 = (1,1,0)$ and $\mathbf{v}_2 = (0,1,1)$.

**Solution.** By Proposition 1.10.2(3),

$$\mathbf{u} = \mathbf{v}_1 \times \mathbf{v}_2 = (1,-1,1)$$

is a normal vector for the plane. Therefore, an equation for it is

$$(1,-1,1) \bullet ((x,y,z) - (1,0,3)) = 0$$

or

$$x - y + z = 4.$$

## 1.11 The Generalized Inverse Matrix

Let

$$T: \mathbf{R}^m \to \mathbf{R}^n$$

be a linear transformation. Now normally one would not expect this arbitrary map T to have an inverse, especially if m > n, but it turns out that it is possible to define something close to that that is useful. Define a map

$$T^+ : \mathbf{R}^n \to \mathbf{R}^m$$

as follows: See Figure 1.19. Let $\mathbf{b} \in \mathbf{R}^n$. The point $\mathbf{b}$ may not be in the image of T, im(T), since we are not assuming that T is onto, but im(T) is a plane in $\mathbf{R}^n$. Therefore, there is a unique point $\mathbf{c} \in \mathrm{im}(T)$ that is closest to $\mathbf{b}$ (Theorem 4.5.12). If the transformation T is onto, then obviously $\mathbf{c} = \mathbf{b}$. It is easy to show that $T^{-1}(\mathbf{c})$ is a plane in $\mathbf{R}^m$ that is parallel to the kernel of T, ker(T). This plane will meet the orthogonal complement of the kernel of T, $\ker(T)^\perp$, in a unique point $\mathbf{a}$. For an alternative definition of the point $\mathbf{a}$ write $\mathbf{R}^m$ in the form

$$\mathbf{R}^m = \ker(T) \oplus \ker(T)^\perp$$

and let

$$\varphi = T \mid \ker(T)^\perp : \ker(T)^\perp \to \mathrm{im}(T).$$

It is easy to show that $\varphi$ is an isomorphism and $\mathbf{a} = \varphi^{-1}(\mathbf{c})$. In either case, we define $T^+(\mathbf{b}) = \mathbf{a}$.

**Definition.** The map $T^+$ is called the *generalized* or *Moore-Penrose inverse* of T.

**1.11.1. Lemma.** $T^+$ is a well-defined linear transformation.

**Proof.** Easy.

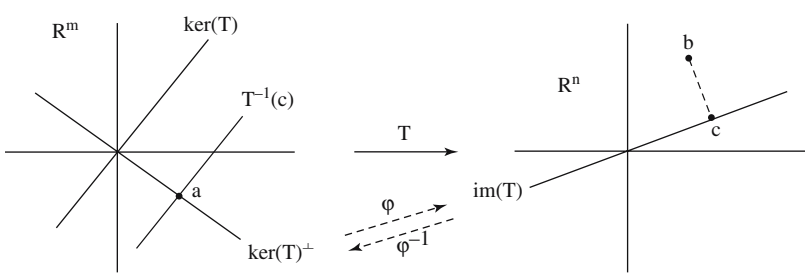



**Figure 1.19.** The geometry behind the generalized inverse.

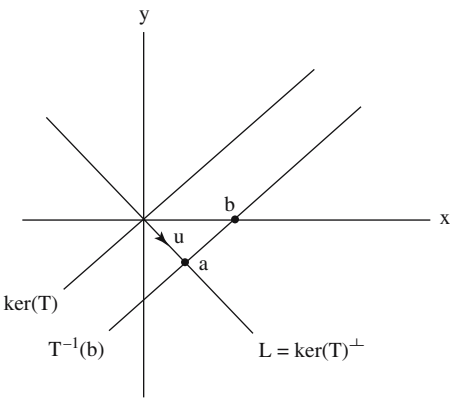



**Figure 1.20.** Computing the generalized inverse for Example 1.11.2.

**1.11.2. Example.** Consider the map $T:\mathbf{R}^2 \to \mathbf{R}$ defined by $T(x,y) = x - y$. Let $b \in \mathbf{R}$. We want to show that the generalized inverse $T^+:\mathbf{R} \to \mathbf{R}^2$ is defined by

$$T^+(b) = \frac{b}{2}(1,-1).$$

**Solution.** See Figure 1.20. The kernel of T, ker(T), is the line $x = y$ in $\mathbf{R}^2$. The orthogonal complement of ker(T) is the line **L** defined by $x + y = 0$. If $\mathbf{a} = T^+(b)$, then **a** is the point where the line $T^{-1}(b)$ meets **L**. Clearly, such a point **a** is just the orthogonal projection of the vector (b,0) on **L**, that is,

$$\mathbf{a} = (\mathbf{u} \bullet (b,0))\,\mathbf{u} = \frac{b}{2}(1,-1),$$

for any unit direction vector **u** for **L** (Theorems 4.5.12 and 1.4.6). For example, we could choose

$$\mathbf{u} = \frac{1}{\sqrt{2}}(1,-1).$$

Of interest to us is the matrix version of the generalized inverse. Let A be an arbitrary real $m \times n$ matrix. Let $T:\mathbf{R}^m \to \mathbf{R}^n$ be the natural linear transformation associated to this matrix by the formula $T(\mathbf{x}) = \mathbf{x}A$.

**Definition.** The $n \times m$ matrix $A^+$ for the generalized inverse $T^+$ is called the *generalized inverse* or *pseudo-inverse* or *Moore-Penrose inverse matrix* for A.

**1.11.3. Theorem.**

(1) The generalized inverse matrix $A^+$ for a matrix A satisfies

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad \left(A^+A\right)^T = A^+A, \quad \text{and} \quad \left(AA^+\right)^T = AA^+.$$

(2) The generalized inverse matrix is **uniquely** defined by the identities in (1), that is, if G is a matrix satisfying

$$AGA = A, \quad GAG = G, \quad (GA)^T = GA, \quad \text{and} \quad (AG)^T = AG,$$

then $G = A^+$.

**Proof.** See [Penr55] or [RaoM71].

**1.11.4. Corollary.**

(1) If A is a real $m \times n$ matrix of rank n, then $A^+ = (A^TA)^{-1}A^T$.
(2) If A is a real $m \times n$ matrix of rank m, then $A^+ = A^T(AA^T)^{-1}$.

**Proof.** The Corollary follows from Theorem 1.11.3(2). For part (1), it is easy to check that $A^TA$ is a nonsingular $n \times n$ matrix and $(A^TA)^{-1}A^T$ satisfies the stated identities. Part (2) follows from a similar argument.

**1.11.5. Example.** To compute the matrix $A^+$ for the map $T^+$ in Example 1.11.2 above.

**Solution.** In this case, we have that $A^T = (1\ -1)$, so that $A^TA = 2$ and

$$A^+ = \left(A^TA\right)^{-1}A^T = \frac{1}{2}(1\ \ -1),$$

which agrees with our formula for $T^+$.

A nice application of the generalized inverse matrix and Corollary 1.11.4 is to a linear least squares approximation problem. Suppose that we are given a real $m \times n$ matrix A with $n > m$ and $\mathbf{b} \in \mathbf{R}^n$. We want to solve the equation

$$\mathbf{x}\,A = \mathbf{b} \tag{1.27}$$

for $\mathbf{x} \in \mathbf{R}^m$. Unfortunately, the system of equations defined by (1.27) is overdetermined and may not have a solution. The best that we can do in general is to solve the following problem:

**A linear least squares approximation problem:** Given an $m \times n$ matrix A with $n > m$ and $\mathbf{b} \in \mathbf{R}^n$, find a point $\mathbf{a}_0 \in \mathbf{R}^m$ that minimizes the distances $|\mathbf{a}A - \mathbf{b}|$, that is, find $\mathbf{a}_0$ so that

$$|\mathbf{a}_0A - b| = \min_{\mathbf{a}\in\mathbf{R}^m}\{\ |\mathbf{a}A - \mathbf{b}|\ \}. \tag{1.28}$$

It is easy to explain the name of the problem. Let

$$A = (a_{ij}), \quad \mathbf{a} = (a_1, a_2, \ldots, a_m), \quad \text{and} \quad \mathbf{b} = (b_1, b_2, \ldots, b_n).$$

Equation (1.28) is trying to find the minimum for

$$\sqrt{\sum_{i=1}^{n}\left[(a_1a_{1i}+a_2a_{2i}+\ldots+a_ma_{mi})-b_i\right]^2}.$$

To put it another way, we are trying to find the m-plane $\mathbf{X}$ in $\mathbf{R}^{m+1}$ through the origin defined by an equation of the form

$$a_1x_1+a_2x_2+\ldots+a_mx_m-x_{m+1}=0 \tag{1.29}$$

that best fits the data points $\mathbf{p}_i = (a_{1i},a_{2i},\ldots,a_{mi},b_i) \in \mathbf{R}^{m+1}$ in the sense that the sum of the distances of the points $\mathbf{p}_i$ to $\mathbf{X}$ is a minimum.

**1.11.6. Theorem.** If the matrix A in the linear least squares approximation problem above has rank m, then there is a unique solution $\mathbf{a}_0$ defined by

$$\mathbf{a}_0 = A^+(\mathbf{b}) = A^T\left(AA^T\right)^{-1}(\mathbf{b}).$$

**Proof.** This is clear from the definition of the generalized inverse and Corollary 1.11.4(2). The uniqueness follows from the fact that the kernel of the linear transformation associated to A is $\mathbf{0}$.

We need to point out that the planes defined by equation (1.29) are a subset of all the m-planes through the origin, so that our particular approximation problem had a bias built into it. Here is the usual statement of the unbiased general problem. One uses squares of the distances to avoid having to deal with square roots. The minimization problem has the same answer in either case.

> **The linear least squares approximation problem:** Given a set of points $\mathbf{p}_i$ in $\mathbf{R}^{m+1}$ find the m-plane $\mathbf{X}$ in $\mathbf{R}^{m+1}$ with the property that the sum of the squares of the distances of the points $\mathbf{p}_i$ to $\mathbf{X}$ is a minimum.

Because of the bias in the allowed solution to our approximation problem, Theorem 1.11.6 does not always solve the general problem. For example, consider the points (–1,1), (–1,2), (–1,3), (1,1), (1,2), and (1,3) in $\mathbf{R}^2$. The line that best approximates this data is clearly the vertical line x = 0. Theorem 1.11.6 would give us simply the point (0,0). The reason for this is that the vertical line does not have an equation of the form (1.29). Of course, Theorem 1.11.6 does give the expected answer "most" of the time but one must make sure that this answer does not lie in the set of planes excluded by equation (1.29).

There is another special case where Theorem 1.11.6 does not give a satisfactory answer, namely, in the case where $\mathbf{b}$ is zero and we have a homogeneous equation

$$\mathbf{x}\,A = \mathbf{0}. \tag{1.30}$$

A homogeneous equation like (1.30) always has a solution $\mathbf{x} = \mathbf{0}$. This is what Theorem 1.11.6 would give us. Of course, this is the uninteresting solution and we are probably looking for a nonzero solution. We will be able to use Theorem 1.11.6 if we rewrite

things. Suppose that we look for a solution $\mathbf{a} = (a_1, a_2, \ldots, a_m)$ to equation (1.30) with $a_m \neq 0$. Then equation (1.30) can be written in the form

$$\begin{pmatrix} \frac{a_1}{a_m} & \cdots & \frac{a_{m-1}}{a_m} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{m-1,1} & \cdots & a_{m-1,n} \end{pmatrix} = -(a_{m1} \quad \cdots \quad a_{mn}).$$

There is actually no loss of generality in assuming that $a_m = 1$. This equation is again of the form (1.27) and if we assume that the right-hand side of this equation is not the zero vector, then we can again apply Theorem 1.11.6 and get what we want in the same sense as before. The only problem however is that we could assume different coordinates of $\mathbf{a}$ to be nonzero and for each choice we shall get different solutions. The general point to remember then is that the approach to the linear least squares problem that we described above works well but the answer that we get depends on the assumptions that we make.

We finish this section with two results about decompositions of matrices.

**1.11.7. Theorem.** Let A be a real $m \times n$ matrix of rank r. Then there exists an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and a diagonal $m \times n$ matrix

$$D = \begin{pmatrix} \sigma_1 & & 0 & \mathbf{0} \\ & \ddots & & \\ 0 & & \sigma_r & \\ \mathbf{0} & & & \mathbf{0} \end{pmatrix},$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, so that

$$A = UDV^T. \tag{1.31}$$

**Proof.** See [ForM67] or [RaoM71].

**Definition.** The decomposition of A in equation (1.31) is called the *singular value decomposition* of A. The $\sigma$s are called the *singular values* of A.

The singular value decomposition of a matrix has useful applications. One interpretation of Theorem 1.11.7 is that, up to change of coordinates, every linear transformation $T: \mathbf{R}^m \to \mathbf{R}^n$ of rank r has the form $T(\mathbf{e}_i) = \sigma_i \mathbf{e}_i$, $1 \leq i \leq r$. More precisely, one can find orthonormal bases $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ of $\mathbf{R}^m$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ of $\mathbf{R}^n$, so that $T(\mathbf{u}_i) = \sigma_i \mathbf{v}_i$, $1 \leq i \leq r$.

**1.11.8. Theorem.** Let A be a real $m \times n$ matrix of rank r. If A has the singular value decomposition shown in equation (1.31), then

$$A^+ = VD^+U^T,$$

and the $n \times m$ matrix $D^+$ is given by

$$D^+ = \begin{pmatrix} 1/\sigma_1 & & 0 & \mathbf{0} \\ & \ddots & & \\ 0 & & 1/\sigma_r & \\ \mathbf{0} & & & \mathbf{0} \end{pmatrix}.$$

**Proof.** One uses Theorem 1.11.3(2) and shows that $VD^+U^T$ satisfies the appropriate identities.

## 1.12 EXERCISES

### Section 1.2

1.2.1. Suppose that the equations

$$ax + by = c \quad \text{and} \quad a'x + b'y = c'$$

define the same line **L.** Show that $a' = ka$, $b' = kb$, and $c' = kc$ for some nonzero real number k.

1.2.2. Show that the equation form and point-direction-vector form of the definition of a line in the plane agree.

1.2.3. Find the equation for all lines in the plane through the point (2,3).

1.2.4. Find the parametric equations of the line through the points (0,1,2) and (–1,–1,–1).

1.2.5. If $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$, show that $[\mathbf{p},\mathbf{q}] = [\mathbf{q},\mathbf{p}]$.

1.2.6. Let $a, b \in \mathbf{R}$ with $a \le b$. Show that the interval [a,b] consists of the same numbers as the segment [a,b] where a and b are thought of as vectors. The difference between a segment and an interval in **R** is that the **interval** [a,b] is defined to be empty if $b < a$, whereas this is not the case for segments. In fact, as **segments** (in $\mathbf{R}^1$) [a,b] = [b,a].

1.2.7. Consider the line **L** through (1,–1,0) with direction vector (–1,–1,2). Find the two points on **L** that are a distance 2 from the point (0,–2,2).

### Section 1.3

1.3.1. Find the cosines of the angles between the following pairs of vectors. Which pairs are perpendicular? Which pairs are parallel?

(a) (3,1), (1,3) (b) (1,2), (–4,2) (c) (1,2), (–4,–8) (d) (–3,0), (2,1)

### Section 1.4

1.4.1. Fill in the missing details in the proof of Theorem 1.4.4.

1.4.2. Prove Theorem 1.4.6.

1.4.3. Find the orthogonal projection of (–1,2,3) on (1,0,1).

1.4.4. Use the Gram-Schmidt algorithm to replace the vectors (1,0,1), (0,1,1), and (2,–3,–1) by an orthonormal set of vectors that spans the same subspace.

1.4.5. This exercise shows that equation (1.14) in Theorem 1.4.6 gives the wrong answer if the vectors $\mathbf{u}_i$ do not form an orthonormal basis. Consider the vector $\mathbf{v} = (1,2,3)$ in $\mathbf{R}^3$. Its orthogonal projection onto $\mathbf{R}^2$ should clearly be (1,2,0). Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be a basis for $\mathbf{R}^2$ and let

$$\mathbf{w} = (\mathbf{v} \bullet \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{v} \bullet \mathbf{u}_2)\mathbf{u}_2.$$

(a) If $\mathbf{u}_1 = (1,0,0)$ and $\mathbf{u}_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$, show that $\mathbf{w} \neq (1,2,0)$.

(b) If $\mathbf{u}_1 = (2,0,0)$ and $\mathbf{u}_2 = (0,3,0)$, show that $\mathbf{w} \neq (1,2,0)$.

**Section 1.5**

1.5.1. Suppose that $\mathbf{X}$ is a k-dimensional plane in $\mathbf{R}^n$ and that

$$\mathbf{X} = \{\mathbf{p} + \mathbf{v} \mid \mathbf{v} \in \mathbf{V}\} = \{\mathbf{q} + \mathbf{w} \mid \mathbf{w} \in \mathbf{W}\},$$

where $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$ and $\mathbf{V}$ and $\mathbf{W}$ are k-dimensional vector subspaces of $\mathbf{R}^n$. Show that $\mathbf{V} = \mathbf{W}$.

1.5.2. Fill in the missing details in the proof of Proposition 1.5.2.

1.5.3. Prove that the intersection of two planes is a plane.

1.5.4. Prove that if $\mathbf{X}$ is a plane, then aff($\mathbf{X}$) = $\mathbf{X}$.

1.5.5. Prove Lemma 1.5.6.

1.5.6. Prove Lemma 1.5.7.

1.5.7. (a) Prove that two lines in $\mathbf{R}^2$ are parallel if and only if they have parallel direction vectors.

(b) Let $\mathbf{L}$ and $\mathbf{L}'$ be lines in $\mathbf{R}^2$ defined by the equations $ax + by = c$ and $a'x + b'y = c'$, respectively. Prove that $\mathbf{L}$ and $\mathbf{L}'$ are parallel if and only if $a' = ka$ and $b' = kb$ for some nonzero constant k.

1.5.8. Prove Theorem 1.5.10.

1.5.9. Find a basis for the plane $x - 3y + 2z = 12$ in $\mathbf{R}^3$.

1.5.10. Find the equation of all planes in $\mathbf{R}^3$ that are orthogonal to the vector (1,2,3).

1.5.11. Find the equation of the plane containing the points (1,0,1), (3,–1,1), and (0,1,1).

1.5.12. Find the equation for the plane in $\mathbf{R}^3$ that contains the point (1,2,1) and is parallel to the plane defined by $x - y - z = 7$.

1.5.13. Find an equation for all planes in $\mathbf{R}^3$ that contain the point (1,2,1) and are orthogonal to the plane defined by $x - y - z = 7$.

1.5.14. Find an orthonormal basis for the plane $x + 2y - z = 3$.

1.5.15. Let $\mathbf{X}$ be the plane defined by $2x + y - 3z = 7$. Let $\mathbf{v} = (2,1,0)$.

(a) Find the orthogonal projection of $\mathbf{v}$ on $\mathbf{X}$.

(b) Find the orthogonal complement of $\mathbf{v}$ with respect to $\mathbf{X}$.

1.5.16. Find the point-normals equation for the line

$$\begin{aligned} x &= 1 - 3t \\ y &= 2t \\ z &= 3 + t \end{aligned}$$

1.5.17. Determine whether the halfplanes $2y - x \geq 0$, $y - 2x + 2 \geq 0$, and $-4y + 2x + 4 \leq 0$ have a nonempty intersection or not.

1.5.18. Let $\mathbf{V}$ and $\mathbf{W}$ be subspaces of $\mathbf{R}^n$ of dimension s and t, respectively. Assume that $s + t \geq n$. Prove that $\mathbf{V}$ is transverse to $\mathbf{W}$ if and only if one of the following holds:

(a) $\mathbf{V} + \mathbf{W} = \mathbf{R}^n$.
(b) If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_t$ are bases for $\mathbf{V}$ and $\mathbf{W}$, respectively, then the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_t$ span $\mathbf{R}^n$.

1.5.19. **Definition.** Any subset $\mathbf{X}$ in $\mathbf{C}^n$ of the form

$$\mathbf{X} = \{\mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \ldots + t_k\mathbf{v}_k \mid t_1, t_2, \ldots, t_k \in \mathbf{C}\},$$

where $\mathbf{p}$ is a fixed point and the $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are fixed linearly independent vectors in $\mathbf{C}^n$ is called a *complex k-dimensional plane* (through $\mathbf{p}$). If $k = 1$, then $\mathbf{X}$ is called a *complex line*.

(a) Prove that a complex line in $\mathbf{C}^2$ can also be expressed as the set of points $(x,y) \in \mathbf{C}^2$ satisfying an equation of the form

$$ax + by = c,$$

for fixed $a, b, c \in \mathbf{C}$ with $(a,b) \neq (0,0)$.
(b) Prove that the real points of a complex plane in $\mathbf{C}^n$ lie on a plane in $\mathbf{R}^n$.

## Section 1.6

1.6.1. Prove Lemma 1.6.1.

1.6.2. Determine whether the following pairs of ordered bases of $\mathbf{R}^2$ determine the same orientation:

(a) $((1,-2), (-3,2))$ and $((1,0), (-2,3))$
(b) $((-1,1), (1,2))$ and $((1,-2), (1,-4))$

Solve this exercise in two ways: First, use only the definition of orientation and then check your answer using the matrix approach of Lemma 1.6.4.

1.6.3. Why is "Does $((1,-2), (-2,4))$ induce the standard orientation of the plane?" a meaningless question?

1.6.4. (a) Find a vector $(a,b)$ so that the basis $((-2,-3), (a,b))$ determines the standard orientation of the plane.
(b) Find a vector $(a,b,c)$ so that the basis $((2,-1,0), (-2,-1,0), (a,b,c))$ determines the standard orientation of 3-space.

1.6.5. Show that the angle between oriented hyperplanes $(\mathbf{X},\sigma)$ and $(\mathbf{Y},\tau)$ is well defined. Specifically, show that it does not depend on the choice of the normal vectors $\mathbf{v}_n$ and $\mathbf{w}_n$ for $\mathbf{X}$ and $\mathbf{Y}$, in the definition.

1.6.6. Let $\mathbf{L}$ be an oriented line and let $\mathbf{p}$ and $\mathbf{q}$ be two points on $\mathbf{L}$. Prove that

$$\begin{aligned}\|\mathbf{pq}\| &= 0, && \text{if } \mathbf{p} = \mathbf{q},\\ &= |\mathbf{pq}|, && \text{if the vector } \mathbf{pq} \text{ induces the same orientation on } \mathbf{L}, \text{ and}\\ &= -|\mathbf{pq}|, && \text{if } \mathbf{pq} \text{ induces the opposite orientation on L.}\end{aligned}$$

1.6.7. Let $(\mathbf{V},\sigma)$ and $(\mathbf{W},\tau)$ be two oriented n-dimensional vector spaces and let $T:\mathbf{V} \to \mathbf{W}$ be a nonsingular linear transformation. Show that T is orientation preserving if

$$\tau = [T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)]$$

for any **one** ordered bases $(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n)$ of $\mathbf{V}$ with the property that $\sigma = [\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_n]$.

**Section 1.7**

1.7.1. Show that each halfplane in $\mathbf{R}^n$ is convex.

1.7.2. Show that if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ are convex sets, then their intersection is convex.

1.7.3. If $\mathbf{X}$ is convex, show that $\text{conv}(\mathbf{X}) = \mathbf{X}$.

1.7.4. Show that $\text{conv}(\{\mathbf{p}_0,\mathbf{p}_1\}) = [\mathbf{p}_0,\mathbf{p}_1]$.

1.7.5. Let $\boldsymbol{\sigma}$ be the two-dimensional simplex defined by the vertices $\mathbf{v}_0 = (-2,-1)$, $\mathbf{v}_1 = (3,0)$, and $\mathbf{v}_2 = (0,2)$. The points of $\boldsymbol{\sigma}$ can be described either with Cartesian or barycentric coordinates (with respect to the vertices listed in the order given above).

(a) Find the Cartesian coordinates of the point $\mathbf{p}$ whose barycentric coordinates are $\left(\frac{1}{4},\frac{5}{12},\frac{1}{3}\right)$.

(b) Find the barycentric coordinates of the point $\mathbf{q}$ whose Cartesian coordinates are (0,0).

1.7.6. Show that the simplicial map from the 1-simplex [2,5] to the 1-simplex [3,7] that sends 2 to 3 and 5 to 7 agrees with the "standard" linear map between the intervals, namely,

$$g(x) = \frac{4}{3}x + \frac{1}{3}.$$

1.7.7. Generalize Exercise 1.7.6 and show that the simplicial map from [a,b] to [c,d] agrees with the standard linear map.

**Section 1.8**

1.8.1. Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}.$$

Find a matrix P so that $P^{-1}AP$ is a diagonal matrix.

**Section 1.9**

1.9.1. Let

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 5 & -4 \\ -3 & -4 & 8 \end{pmatrix}.$$

Find a nonsingular matrix C so that $CAC^T$ is a diagonal matrix.

**Section 1.10**

1.10.1. Prove Proposition 1.10.3. (Hint: First show that, if $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are the standard basis vectors in $\mathbf{R}^3$, then $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$, $\mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2$, and $\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$.)

1.10.2. Prove Proposition 1.10.4.

**Note:** The properties will not be hard to prove if one uses the definition and basic properties of determinants. This shows once again how valuable a good definition is because some textbooks, especially in the physical sciences, deal with cross products in very messy ways. Although it is our intuition which leads us to useful concepts, it is usually a good idea not to stop with the initial insight but probe a little further and really capture their essence.

1.10.3. Prove that if $\mathbf{u}, \mathbf{v} \in \mathbf{R}^3$ are orthogonal unit vectors, then $(\mathbf{u} \times \mathbf{v}) \times \mathbf{u} = \mathbf{u}$.

1.10.4. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^3$. Prove

(a) $\mathbf{u} \bullet (\mathbf{v} \times \mathbf{w}) = \det \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix}$.

(b) $\mathbf{u} \bullet (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \bullet (\mathbf{w} \times \mathbf{u}) = \mathbf{w} \bullet (\mathbf{u} \times \mathbf{v})$.

(The quantity $\mathbf{u} \bullet (\mathbf{v} \times \mathbf{w})$ is called the *triple product* of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$.)

CHAPTER 2

# Affine Geometry

## 2.1 Overview

The next two chapters deal with the analytic and geometric properties of some important transformations of $\mathbf{R}^n$. This chapter discusses the group of affine maps and its two important subgroups, the group of similarities and the group of motions. Affine maps are the transformations that preserve parallelism. Similarities are the affine transformations that preserve angles. Motions are the distance-preserving similarities and their study is equivalent to the study of metric properties of Euclidean space. As a historical note, this reduction of geometric problems to algebra (namely the study of certain groups in our case) was initiated by the German mathematician Felix Klein at the end of the $19^{th}$ century.

Except for some definitions and a few basic facts, the first part of the chapter (Sections 2.2–2.4) concentrates on the important special case of the plane $\mathbf{R}^2$. Presenting a lot of details in the planar case where it is easier to draw pictures should make it easier to understand what happens in higher dimensions since the generalizations are, by and large, straightforward.

Motions are probably the most well-known affine maps and we analyze planar motions in quite some detail in Section 2.2. Section 2.2.8 introduces the concept of a frame. Frames are an extremely useful way to deal with motions and changing from one coordinate system to another. It is not an overstatement to say that a person who understands frames will find working with motions a triviality. There is a brief discussion of similarities in Section 2.3 and affine maps in Section 2.4. Parallel projections are defined in Section 2.4.1. Section 2.5 extends the main ideas from the plane to higher dimensions. The important case of motions in $\mathbf{R}^3$ is treated separately in Sections 2.5.1 and 2.5.2.

There is not enough space to prove everything in this chapter and it will be up to the reader to fill in missing details or to look them up in the references. Hopefully, the details we do provide in conjunction with what we did in Chapter 1 will make filling in missing details easy in most cases. Unproved facts are included because it was felt that they were worth knowing about and help as motivation for the next

chapter on projective transformations. References to where proofs may be found are given in those cases where difficult results are stated but not proved.

Finally, we want to emphasize one point. The single most important topic in this chapter is that of frames. Frames are so simple (they are just orthonormal bases), yet if the reader masters their use, then dealing with transformations will be a snap!

## 2.2 Motions

**Definition.** A transformation $M: \mathbf{R}^n \to \mathbf{R}^n$ is called a *motion* or *isometry* or *congruent transformation* of $\mathbf{R}^n$ if

$$|M(\mathbf{p})M(\mathbf{q})| = |\mathbf{pq}|,$$

for every pair of points $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$.

In simple terms, motions are *distance-preserving maps*. If one concentrates on that aspect, then the term "isometry" is the one that mathematicians normally use when talking about distance-preserving maps between arbitrary spaces. The term "motion" is popular in the context of $\mathbf{R}^n$.

**2.2.1. Theorem.**

(1) Motions preserve the betweenness relation.
(2) Motions preserve collinearity and noncollinearity.
(3) Motions send lines to lines.

**Proof.** To prove (1), let M be a motion and let $\mathbf{C}$ be a point between two points $\mathbf{A}$ and $\mathbf{B}$. Let $(\mathbf{A}',\mathbf{C}',\mathbf{B}') = M(\mathbf{A},\mathbf{C},\mathbf{B})$. We must show that $\mathbf{C}'$ is between $\mathbf{A}'$ and $\mathbf{B}'$. Now

$$\begin{aligned} |\mathbf{A}'\mathbf{B}'| &= |\mathbf{AB}| \\ &= |\mathbf{AC}| + |\mathbf{CB}| \\ &= |\mathbf{A}'\mathbf{C}'| + |\mathbf{C}'\mathbf{B}'|. \end{aligned}$$

The first and third equality above follows from the definition of a motion. The second follows from Proposition 1.2.3. Using Proposition 1.2.3 again proves (1). Parts (2) and (3) of the theorem clearly follow from (1).

**2.2.2. Lemma.** Let M be a motion. If

$$\mathbf{C} = \mathbf{A} + t\,\mathbf{AB} = (1-t)\,\mathbf{A} + t\mathbf{B},$$

then

$$M(\mathbf{C}) = M(\mathbf{A}) + tM(\mathbf{A})M(\mathbf{B}) = (1-t)M(\mathbf{A}) + tM(\mathbf{B}).$$

**Proof.** Let $(\mathbf{A}',\mathbf{B}',\mathbf{C}') = M(\mathbf{A},\mathbf{B},\mathbf{C})$. Since M is a motion,

$$|\mathbf{A}'\mathbf{C}'| = |\mathbf{AC}| = |t|\,|\mathbf{AB}| = |t|\,|\mathbf{A}'\mathbf{B}'|.$$

The proof is divided into cases.

**Case 1.** $0 \le t \le 1$.
**Case 2.** $1 < t$.
**Case 3.** $t < 0$.

In Case 1, **C** is between **A** and **B**. By Proposition 1.2.4, only

$$\mathbf{X}_1 = \mathbf{A}' + t\mathbf{A}'\mathbf{B}' \quad \text{or} \quad \mathbf{X}_2 = \mathbf{A}' - t\mathbf{A}'\mathbf{B}'$$

are solutions to the equation

$$|\mathbf{A}'\mathbf{X}'| = |t|\,|\mathbf{A}'\mathbf{B}'|.$$

Of these, only $\mathbf{X}_1$ lies between $\mathbf{A}'$ and $\mathbf{B}'$. By part (1) of Theorem 2.2.1 we have that $\mathbf{C}' = \mathbf{X}_1$, which proves the lemma. The proofs in the other two cases are similar and are left as exercises to the reader. Note that in Case 2 **B** is between **A** and **C** and in Case 3 **A** is between **C** and **B**.

**2.2.3. Lemma.** Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be two distinct lines in the plane which intersect in a point **C**. Let **P** be any point not on either of these lines. Then there exist two distinct points **A** and **B** on $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively, so that **P** lies on the line **L** determined by **A** and **B**.

**Proof.** See Figure 2.1. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be direction vectors for $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively. These vectors are linearly independent since the lines are not parallel. Let $\mathbf{A} = \mathbf{C} + a\mathbf{v}_1$ be any point on $\mathbf{L}_1$ with $a > 0$ and let **L** be the line determined by **P** and **A**. To find the intersection of **L** and $\mathbf{L}_2$, we must solve the equation

$$\mathbf{P} + s\,\mathbf{PA} = \mathbf{C} + t\,\mathbf{v}_2$$

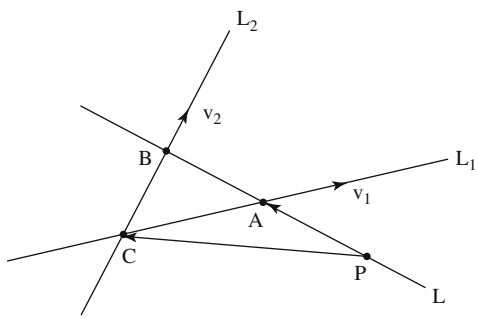



**Figure 2.1.** Proving Lemma 2.2.3.

for real numbers s and t. This equation can be rewritten as

$$\text{sa}\ \mathbf{v}_1 - \text{t}\ \mathbf{v}_2 = (1 - \text{s})\ \mathbf{PC}. \tag{2.1}$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, s cannot be 1 and equation (2.1) has a unique solution for s and t. Let $\mathbf{B} = \mathbf{P} + \text{s}\mathbf{PA}$.

**2.2.4. Lemma.** A motion M is a one-to-one and onto map.

**Proof.** The first part, that M is one-to-one, is easy, because if the distance between the images of two points under M is zero, then so is the distance between the two points by the definition of a motion.

Showing that M is onto is harder and we only prove it in the planar case here. See [Gans69] for the general case. We begin by proving a stronger version of Theorem 2.2.1 (3).

**Claim.** M maps lines **onto** lines.

Let **L** be a line. We already know that M(**L**) is contained in a line **L′**. Let **C′** be any point of **L′**. We must show that there is a point **C** in **L** with M(**C**) = **C′**. To this end, choose any two distinct points **A** and **B** of **L** and let (**A′**,**B′**) = M(**A**,**B**). Then **C′** = **A′** + t**A′B′** for some t. It follows from Lemma 2.2.2 that **C′** = M(**A** + t**AB**) and the claim is proved.

We are ready to prove that planar motions are onto. See Figure 2.2. Let **P′** be any point of $\mathbf{R}^2$. We must show that **P′** = M(**P**) for some point **P**. Take three noncollinear points **A, B,** and **C** and let (**A′**,**B′**,**C′**) = M(**A**,**B**,**C**). Let $\mathbf{L}_1'$ be the line that contains the points **A′** and **B′** and let $\mathbf{L}_2'$ be the line that contains **A′** and **C′**. We just showed that all the points on these two lines are in the image of M. Assume that **P′** is not on these two lines. By Lemma 2.2.3 there are two points **D′** and **E′** on these lines so that **P′** is on the line **L′** determined by **D′** and **E′** and hence in the image of M. The planar case of Lemma 2.2.4 is proved.

Although much of what we shall prove about motions depends only on their distance-preserving property and not on their domain, the domain **can** be important. The following example shows that Lemma 2.2.4 definitely uses the fact that the domain of the motion is all of the plane:

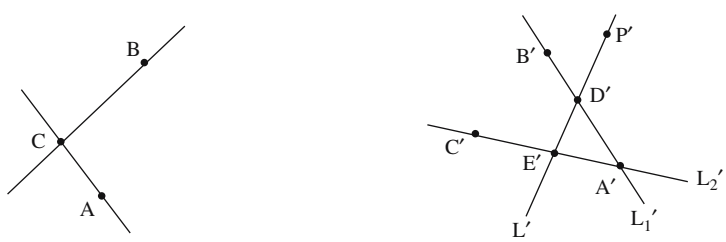



**Figure 2.2.** Proving motions are onto maps.

**2.2.5. Example.** Let $\mathbf{X} = \{ (x,y) \mid x > 0 \} \subseteq \mathbf{R}^2$ and define the distance-preserving map $T:\mathbf{X} \to \mathbf{X}$ by $T(x,y) = (x + 1,y)$. The map T is clearly not onto.

**2.2.6. Theorem.** Motions form a group under composition.

**Proof.** Exercise.

The idea of a motion as a distance-preserving map is intuitively simple to understand, but it is not very useful for making computations. In the process of deriving a simple analytical description of motions, we shall not only get a lot of geometric insights but also get practice in using linear algebra to solve geometric problems. We begin our study of motions with an approach that is used time and again in mathematics. Namely, if faced with the problem of classifying a set of objects, first isolate as many simple and easy-to-understand elements as possible and then try to show that these elements can be used as building blocks from which all elements of the class can be "generated."

## 2.2.1 Translations

The simplest types of motions are translations.

**Definition.** Any map $T:\mathbf{R}^n \to \mathbf{R}^n$ of the form

$$T(\mathbf{p}) = \mathbf{p} + \mathbf{v}, \tag{2.2}$$

where $\mathbf{v}$ is a fixed vector, is called a *translation* of $\mathbf{R}^n$. The vector $\mathbf{v}$ is called the *translation vector* of T.

Writing things out in terms of coordinates, it is easy to see that a map $T(x_1,x_2, \ldots ,x_n) = (x_1',x_2', \ldots ,x_n')$ is a translation if and only if it is defined by equations of the form

$$\begin{aligned} x_1' &= x_1 + c_1 \\ x_2' &= x_2 + c_2 \\ &\cdot \quad \cdot \quad \cdot \\ x_n' &= x_n + c_n, \end{aligned} \tag{2.3}$$

where the $c_i$ are fixed real numbers. Clearly, $(c_1,c_2, \ldots ,c_n)$ is the translation vector of T in this case.

**2.2.1.1. Theorem.** Translations are motions.

**Proof.** This is a simple exercise for the reader.

Here are several simple interesting properties of translations.

**2.2.1.2. Proposition.** A translation T with **nonzero** translation vector $\mathbf{v}$ satisfies the following properties:

(1) T has no fixed points.
(2) T takes lines to lines with the same direction vector (or slope, in the case of the plane).
(3) The only lines fixed by T are those with direction vector **v**. In the case of the plane, the only lines fixed by T are those whose slope is the same as the slope of one of their direction vectors.

**Proof.** (1) and (2) are left as exercises for the reader. To prove (3), consider a line **L** through a point $\mathbf{p}_0$ with direction vector **w**. If T fixes **L**, then T maps a point $\mathbf{p}_0 + t\mathbf{w}$ on **L** to another point on **L** that will have the form $\mathbf{p}_0 + s\mathbf{w}$. Therefore,

$$\begin{aligned}\mathbf{p}_0 + s\mathbf{w} &= T(\mathbf{p}_0 + t\mathbf{w})\\ &= \mathbf{p}_0 + t\mathbf{w} + \mathbf{v},\end{aligned}$$

and so **w** is a multiple of **v**. The converse is just as easy.

In case of the plane, assume that the line **L** fixed by T is defined by the equation

$$ax + by = c. \tag{2.4}$$

The line L has slope –a/b (the case of a vertical line where b is zero is left as an exercise for the reader). If **v** = (h,k), then the slope of **v** is k/h. Choose a point (x,y) on **L**. Since T(x,y) = (x + h,y + k) is assumed to lie on **L**, that point must also satisfy equation (2.4), that is,

$$a(x + h) + b(y + k) = c.$$

Using the identity (2.4) in this last equation implies that ah + bk = 0. This shows that k/h = –a/b and we are done.

### 2.2.2 Rotations in the Plane

Another intuitively simple motion is a rotation of the plane.

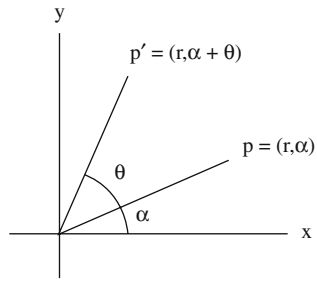
**Definition.** Let $\theta \in \mathbf{R}$. A map $R: \mathbf{R}^2 \to \mathbf{R}^2$ of the form $R(r,\alpha) = (r,\alpha+\theta)$, where points have been expressed in polar coordinates, is called a *rotation about the origin through an angle θ*.

See Figure 2.3. Using polar coordinates was an easy way to define rotations about the origin, but is not convenient from a computational point of view. To derive the equations for a rotation R in Cartesian coordinates, we use the basic correspondence between the polar coordinates (r,α) and Cartesian coordinates (x,y) for a point **p**:

$$\begin{aligned}x &= r\cos\alpha\\ y &= r\sin\alpha\end{aligned} \tag{2.5}$$

Let R(x,y) = (x′,y′). Since R(r,α) = (r,α+θ),

**Figure 2.3.** Defining a rotation with polar coordinates.




$$x' = r\cos(\alpha+\theta) = r\cos\alpha\cos\theta - r\sin\alpha\sin\theta$$
$$y' = r\sin(\alpha+\theta) = r\cos\alpha\sin\theta + r\sin\alpha\cos\theta. \tag{2.6}$$

Substituting (2.5) into (2.6) leads to

**2.2.2.1. Theorem.** The equations for a rotation R about the origin through an angle θ are

$$x' = x\cos\theta - y\sin\theta$$
$$y' = x\sin\theta + y\cos\theta \tag{2.7}$$

In particular, such a rotation is a linear transformation with matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{2.8}$$

**2.2.2.2. Theorem.** Rotations about the origin are motions.

**Proof.** This is proved by direct computations using the definition of a motion and Theorem 2.2.2.1.

**2.2.2.3. Example.** The equations for the rotation R through an angle π/3 are

$$x' = \frac{1}{2}x - \frac{\sqrt{3}}{2}y$$
$$y' = \frac{\sqrt{3}}{2}x + \frac{1}{2}y.$$

Furthermore, notice that the inverse of a rotation through an angle θ is just the rotation through the angle –θ, so that given a rotation it is easy to write down the equations for the inverse. In our example the equations for the inverse are

$$x = \frac{1}{2}x' + \frac{\sqrt{3}}{2}y'$$

$$y = -\frac{\sqrt{3}}{2}x' + \frac{1}{2}y'.$$

We did not have to solve the first set of equations for x and y directly.

**2.2.2.4. Example.** Continuing Example 2.2.2.3, suppose that we would like to find the image **L′** of the line **L** defined by equation $-3x + 2y = 2$.

**Solution.** All we have to do is substitute for x and y:

$$-3\left(\frac{1}{2}x' + \frac{\sqrt{3}}{2}y'\right) + 2\left(-\frac{\sqrt{3}}{2}x' + \frac{1}{2}y'\right) = 2.$$

Simplification of the terms and omitting the "′"on the variables gives that the equation for **L′** is

$$\left(-\frac{3}{2} - \sqrt{3}\right)x + \left(-3\frac{\sqrt{3}}{2} + 1\right)y = 2.$$

Of course, we could also have found two points **p** and **q** on **L** and then computed the equation for the line through the two points R(**p**) and R(**q**), but that would be more work.

So far we have only considered rotations about the origin, but it is easy to define rotations about an arbitrary point.

**Definition.** Let $\mathbf{p} \in \mathbf{R}^2$. The *general rotation R about **p** through an angle* θ is defined by the equation $R = TR_0T^{-1}$, where T is the translation that sends the origin to **p** and $R_0$ is the rotation about the origin through the angle θ. The point **p** is called the *center* of the rotation.

Note that a general rotation is a motion since it is a composite of motions.

**2.2.2.5. Example.** To find the equations for the rotation R about the point (–3,–1) through the angle π/3.

**Solution.** The translation T that sends the origin to (–3,–1) and its inverse $T^{-1}$ are defined by the equations

$$\begin{aligned} T:\ x' &= x - 3 \qquad & T^{-1}:\ x' &= x + 3 \\ y' &= y - 1 & y' &= y + 1 \end{aligned}$$

The equations for the rotation $R_0$ about the origin through the angle π/3 were already computed in Example 2.2.2.3. Therefore, the equations for $R = TR_0T^{-1}$ are

$$x' = \frac{1}{2}(x+3) - \frac{\sqrt{3}}{2}(y+1) - 3$$

$$y' = \frac{\sqrt{3}}{2}(x+3) + \frac{1}{2}(y+1) - 1.$$

The form of the solution to Example 2.2.2.5 generalizes to

**2.2.2.6. Theorem.** The equations for a rotation R about a point **p** = (a,b) through an angle θ are

$$x' = (x-a)\cos\theta - (y-b)\sin\theta + a$$
$$y' = (x-a)\sin\theta + (y-b)\cos\theta + b.$$

**Proof.** Exercise.

Three interesting properties of rotations are

**2.2.2.7. Proposition.**

(1) The only fixed point of a rotation that is not the identity map is its center.
(2) All rotations change the slope of a line unless the rotation is through an angle of 0 or π.
(3) Only the rotations through an angle of 0 or π have a fixed line.

**Proof.** We shall only give a proof of (2). The proof of (1) is left as an exercise and (3) is an immediate consequence of (2).

We already know from Proposition 2.2.1.2 that translations do not change slopes. Therefore it suffices to prove (2) for rotations R about the origin. Let **L** be a line defined by the equation ax + by = c. If R is a rotation through an angle θ and **L′** = R(**L**), then substituting for x and y using the equations for $R^{-1}$ we get that

$$a(x\cos\theta + y\sin\theta) + b(-x\sin\theta + y\cos\theta) = c$$

is an equation for **L′**. The proof of (2) in the special case where either **L** or **L′** is vertical is easy and is left as an exercise. In the rest of the discussion we assume that slopes are defined. It follows that the slope for **L′** is

$$\frac{b\sin\theta - a\cos\theta}{a\sin\theta + b\cos\theta}.$$

But this quotient can never equal the slope of **L** which is –a/b unless sin θ = 0, that is, θ = 0 or π. (Simply set the two expressions equal and simplify the resulting equation to get $b^2 \sin\theta = -a^2 \sin\theta$.) This proves the result.

### 2.2.3 Reflections in the Plane

Another important type of motion is a reflection. Such a motion can be defined in several ways. After giving our definition we shall discuss some of these other characterizations.

**Definition.** Let **L** be a line in the plane. Define a map $S: \mathbf{R}^2 \to \mathbf{R}^2$, called the *reflection about the line* ***L***, as follows: Choose a point **A** on **L** and a **unit** normal vector **N** for **L**. If **P** is any point in $\mathbf{R}^2$, then

$$S(\mathbf{P}) = \mathbf{P}' = \mathbf{P} + 2(\mathbf{PA} \bullet \mathbf{N})\mathbf{N}. \tag{2.9}$$

The line **L** is called the *axis* for the reflection S.

The reader will find Figure 2.4 helpful as we discuss the geometry behind reflections. First, note that $\mathbf{W} = (\mathbf{PA} \bullet \mathbf{N})\mathbf{N}$ is just the orthogonal projection of the vector **PA** onto **N**. Define a point **Q** by the equation

$$\mathbf{PQ} = \mathbf{W} = (\mathbf{PA} \bullet \mathbf{N})\mathbf{N}.$$

Intuitively, it should be clear that **Q** is the point on **L** as shown in Figure 2.4. This does not follow from the definition however and must be proved. The following string of equalities:

$$\mathbf{AQ} \bullet \mathbf{N} = (\mathbf{PQ} + \mathbf{AP}) \bullet \mathbf{N} = [(\mathbf{PA} \bullet \mathbf{N})\mathbf{N} + \mathbf{AP}] \bullet \mathbf{N} = \mathbf{PA} \bullet \mathbf{N} + \mathbf{AP} \bullet \mathbf{N} = 0$$

shows that **Q** satisfies the point-normal form of the equation $\mathbf{AX} \bullet \mathbf{N} = 0$ for the points **X** on the line (or hyperplane) **L**, so that **Q** does indeed lie on **L**. Furthermore, it is easy to check that **AQ** is the orthogonal projection of **AP** on **L**. This means that, if **V** is a unit direction vector for **L**, then $\mathbf{AQ} = (\mathbf{AP} \bullet \mathbf{V})\mathbf{V}$ and we could have defined the reflection S by

$$S(\mathbf{P}) = \mathbf{P} + 2(\mathbf{PA} + \mathbf{AQ}). \tag{2.10}$$

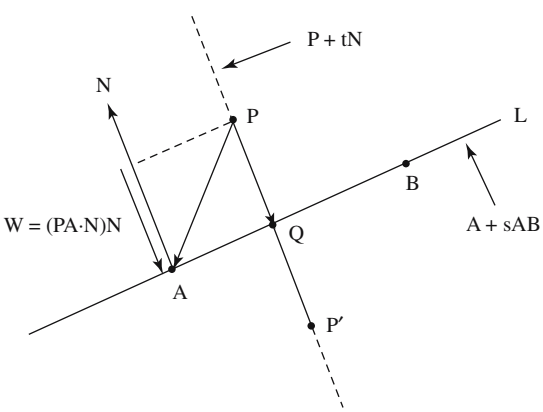



**Figure 2.4.** Defining a reflection in the plane.

This definition has the advantage that one does not need to know a normal vector for the line (only a direction vector or a second point **B** on the line). Of course, finding a normal vector to a line in the plane is trivial. On the other hand, our normal vector definition of a reflection will generalize to higher dimensions later.

Finally, since

$$\mathbf{Q} = \mathbf{P} + (\mathbf{PA} \bullet \mathbf{N})\mathbf{N},$$

we see that **Q** is the point where the line through **P** that is orthogonal to **L** meets **L**. Therefore, another definition of S(**P**) is that we solve for that point **Q** and then define

$$S(\mathbf{P}) = \mathbf{P} + 2\mathbf{PQ}. \tag{2.11}$$

To put it another way, the segment **PP′** is perpendicular to the line **L** and intersects the line at its midpoint **Q**.

**2.2.3.1. Theorem.** Let S be the reflection about a line **L**.

(1) The definition of S depends only on **L** and not on the point **A** and the normal vector **N** that are chosen in the definition. The three definitions of a reflection specified by equations (2.9), (2.10), and (2.11) are equivalent.
(2) If t is chosen so that **P** + t**N** is the point where the line through **P** with direction vector **N** meets the line **L**, then S(**P**) = **P** + 2t**N**.
(3) The fixed points of S are just the points on its axis **L**.
(4) If **L** is the axis of a reflection S and **L′** is a line orthogonal to **L**, then S(**L′**) = **L′**.
(5) Reflections are motions.

**Proof.** Exercise.

**2.2.3.2. Example.** To find the reflection $S_x$ about the x-axis.

**Solution.** If we choose **A** = (0,0) and **N** = (0,1), then **PA** = –**P** and

$$S_x(\mathbf{P}) = \mathbf{P} + 2[(-\mathbf{P}) \bullet (0,1)]\,(0,1),$$

or

$$S_x(x, y) = (x, -y).$$

In other words, $S_x$ has equations

$$\begin{aligned} x' &= x \\ y' &= -y. \end{aligned} \tag{2.12}$$

**2.2.3.3. Example.** To find the reflection S about the line **L** defined by the equation 2x – y + 2 = 0.

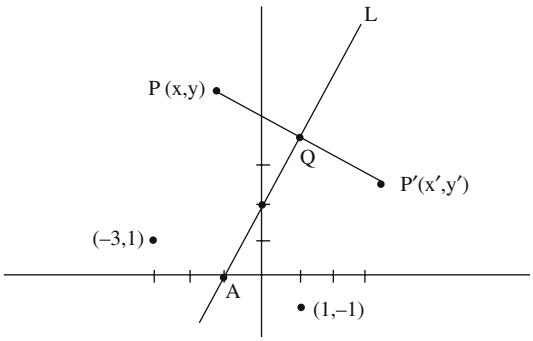



**Figure 2.5.** Example 2.2.3.3.

**Solution.** Let **A** = (–1,0), **N** = $(1/\sqrt{5})(2,-1)$, and let **P**, **Q**, **P′** be as shown in Figure 2.5. Then **PA** = (–x–1,–y). Using the formulas in the definition of a reflection, it follows that

$$\begin{aligned}\mathbf{PQ} &= \left[(-x,-1,-y)\bullet\frac{1}{\sqrt{5}}(2,-1)\right]\frac{1}{\sqrt{5}}(2,-1)\\ &= \left(-\frac{4}{5}x+\frac{2}{5}y-\frac{4}{5},\frac{2}{5}x-\frac{1}{5}y+\frac{2}{5}\right).\end{aligned}$$

Since S(**P**) = **P** + 2**PQ** , we get that the equations for S are

$$\begin{aligned}x' &= -\frac{3}{5}x+\frac{4}{5}y-\frac{8}{5}\\ y' &= \frac{4}{5}x+\frac{3}{5}y+\frac{4}{5}.\end{aligned} \tag{2.13}$$

To check our answer note that S(–3,1) evaluates to (1,–1), which is what it should be. Again see Figure 2.5. Our equations also give that S(**A**) = **A** and S(**B**) = **B**.

**2.2.3.4. Proposition.** If S is the reflection about the line **L** defined by the equation ax + by + c = 0, then

$$S(x,y) = (x,y) + 2\frac{-ax-by-c}{a^2+b^2}(a,b). \tag{2.14}$$

**Proof.** The proof of this formula is based on Theorem 2.2.3.1(2). We know that **N** = (a,b) is a normal vector for **L** (although it may not be a unit vector). Therefore, if **P** = (x,y), to find the point **Q** shown in Figure 2.4, we need to find t so that **P** + t**N** lies on **L.** But

$$a(x+ta)+b(y+tb)+c=0$$

implies that

$$t = \frac{-ax - by - c}{a^2 + b^2}.$$

We get our equation by substituting this t into

$$S(\mathbf{P}) = \mathbf{P} + 2\mathbf{PQ} = \mathbf{P} + 2t\mathbf{N}.$$

**2.2.3.5. Example.** We redo Example 2.2.3.3 using equation (2.14).

**Solution.** In this case

$$t = \frac{-2x + y - 2}{5},$$

so that

$$S(x, y) = (x, y) + 2\frac{-2x + y - 2}{5}(2, -1).$$

This equation simplifies to the same equation for S as before.

A final and more systematic way to compute reflections, one that is easier to remember conceptually (given that one understands translations, rotations about the origin, and the reflection about the x-axis), is based on the often useful general principle that complicated problems should be solved by successively reducing them to simpler ones until one arrives at a primitive problem whose solution is known.

**Case 1** (The primitive problem). The equation for the reflection $S_x$ about the x-axis.

This problem was solved in Example 2.2.3.2 above and we got equations (2.12).

**Case 2.** The equation for a reflection about a line through the origin.

This case can be reduced to the Case 1 by first rotating the line to the x-axis, then using the equation from Case 1, and finally rotating back.

**Case 3** (The general case). The equation for a reflection about an arbitrary line.

By translating the line to a line through the origin we can reduce this case to Case 2, find the equation for that case, and then translate back.

The steps outlined in Cases 1–3 lead to the following characterization of a reflection:

**2.2.3.6. Theorem.** Every reflection S in the plane can be expressed in the form

$$S = T^{-1}R^{-1}S_xRT,$$

where T is a translation, R is a rotation about the origin, and $S_x$ is the reflection about the x-axis.

**Proof.** Exercise.

**2.2.3.7. Example.** To find the equation for the reflection S about the line in Example 2.2.3.3 using this approach.

**Solution.** First translate the line **L** to a line **L′** that passes through the origin via the translation

$$\begin{aligned} T:\ x' &= x+1 \\ y' &= y \end{aligned}$$

Next, let R be the rotation about the origin through the angle $-\theta$ defined by

$$\cos\theta = \frac{1}{\sqrt{5}}, \quad 0 \le \theta \le \pi/2.$$

R will rotate **L′** into the x-axis because $\theta$ is the angle that the line **L** makes with the x-axis. The equations for R and $R^{-1}$ are

$$\begin{aligned} R:\ x' &= \frac{1}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y \qquad & R^{-1}:\ x' &= \frac{1}{\sqrt{5}}x - \frac{2}{\sqrt{5}}y \\ y' &= -\frac{2}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y & y' &= \frac{2}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y. \end{aligned}$$

Finally, if $S_x$ is the reflection about the x-axis, then S is just the composite $T^{-1}R^{-1}S_xRT$. Since the equations for all the maps are known, it is now easy to determine the equations for S and they will again turn out to be the same as the ones as equations (2.13).

The reader might wonder at this point why we bothered to describe the solution in Example 2.2.3.7 since it is more complicated than the one in Example 2.2.3.3. In this instance, the method of Example 2.2.3.7 should simply be considered to be a case of trying to give the reader more insight into how to solve a geometric problem. The approach might not be efficient here but will be in other situations. It is important to realize that there are two types of complexity: one, where we dealing with something that is intellectually difficult, and the other, which may take a lot of time but only involves intellectually simple steps. This is the case with the solution in Example 2.2.3.7. Actually, this type of question will probably come up again later on in this chapter because there are usually many ways to solve problems. Any **particular** problem may very well have an extremely elegant solution that a human might find. On the other hand, a computer is not able to deal with problems on a case-to-case basis and needs a systematic approach.

## 2.2.4 Motions Preserve the Dot Product

**2.2.4.1. Theorem.** If M is a motion with the property that $M(\mathbf{0}) = \mathbf{0}$, then M is a linear transformation, that is,

$$M(a\mathbf{u} + b\mathbf{v}) = a\ M(\mathbf{u}) + b\ M(\mathbf{v}),$$

for all vectors **u** and **v** and real numbers a and b.

**Proof.** We shall show that M is a linear transformation in two steps.

**Claim 1.** $M(\mathbf{u} + \mathbf{v}) = M(\mathbf{u}) + M(\mathbf{v})$.

Define a vector **w** by the equation

$$\mathbf{u} + \mathbf{v} = 2\mathbf{w}. \tag{2.15}$$

This equation can be rewritten as

$$\mathbf{w} = \mathbf{u} + \frac{1}{2}(\mathbf{v} - \mathbf{u}). \tag{2.16}$$

See Figure 2.6. Since $M(\mathbf{0}) = \mathbf{0}$ (which implies that $|M(\mathbf{p})| = |\mathbf{p}|$ for any vector **p**), we can use equation (2.15) and Lemma 2.2.2 to conclude that

$$M(\mathbf{u} + \mathbf{v}) = 2M(\mathbf{w}). \tag{2.17}$$

Similarly, equation (2.16) and Lemma 2.2.2 implies that

$$M(\mathbf{w}) = M(\mathbf{u}) + \frac{1}{2}(M(\mathbf{v}) - M(\mathbf{u})). \tag{2.18}$$

Substituting the expression for $M(\mathbf{w})$ in equation (2.18) into equation (2.17) and simplifying the result proves Claim 1.

**Claim 2.** $M(c\mathbf{v}) = cM(\mathbf{v})$, for any real number c.

This follows from Lemma 2.2.2 (let $\mathbf{A} = \mathbf{0}$, $\mathbf{B} = \mathbf{v}$, and $t = c$ in that Lemma). Theorem 2.2.4.1 is proved.

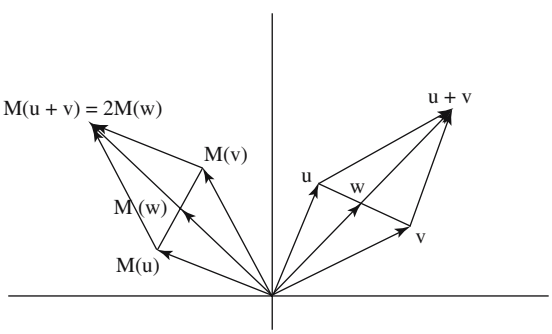



**Figure 2.6.** Proving motions are linear transformations.

**2.2.4.2. Theorem.** Every motion M can be written uniquely in the form $M = T_1M_0 = M_0T_2$, where $T_i$ is a translation and $M_0$ is a motion that fixes the origin, that is, $M_0(\mathbf{0}) = \mathbf{0}$.

**Proof.** Define the translation $T_1$ by $T_1(\mathbf{p}) = \mathbf{p} + M(\mathbf{0})$ and let $M_1 = T_1^{-1}M$. Clearly, $M_1(\mathbf{0}) = \mathbf{0}$ and $M = T_1M_1$. Similarly, if we define the translation $T_2$ by $T_2(\mathbf{p}) = \mathbf{p} - M^{-1}(\mathbf{0})$ and let $M_2 = MT_2^{-1}$, then $M_2(\mathbf{0}) = \mathbf{0}$ and $M = M_2T_2$. Next, we show that $M_2 = M_1$. But by Theorem 2.2.4.1, the motions $M_i$ are linear transformations and so

$$M(\mathbf{p}) = T_1M_1(\mathbf{p}) = M_1(\mathbf{p}) + M(\mathbf{0})$$

and

$$M(\mathbf{p}) = M_2T_2(\mathbf{p}) = M_2(\mathbf{p}) - M_2\big(M^{-1}(\mathbf{0})\big).$$

Therefore, for all $\mathbf{p}$,

$$M_2(\mathbf{p}) - M_2\big(M^{-1}(\mathbf{0})\big) = M_1(\mathbf{p}) + M(\mathbf{0}).$$

The special case where $\mathbf{p}$ is $\mathbf{0}$ shows that $-M_2(M^{-1}(\mathbf{0})) = M(\mathbf{0})$. In other words, we can cancel those terms to get that $M_2(\mathbf{p}) = M_1(\mathbf{p})$. The uniqueness part of the theorem is proved in a similar way.

**2.2.4.3. Lemma.** Let M be a motion and assume that $M(\mathbf{0}) = \mathbf{0}$. Then $M(\mathbf{u}) \bullet M(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v}$ for all vectors $\mathbf{u}$ and $\mathbf{v}$.

**Proof.** The following string of equalities hold because M is a distance preserving map and, by Theorem 2.2.4.1, also a linear transformation:

$$\begin{aligned}
\mathbf{u} \bullet \mathbf{u} + 2\mathbf{u} \bullet \mathbf{v} + \mathbf{v} \bullet \mathbf{v} &= (\mathbf{u} + \mathbf{v}) \bullet (\mathbf{u} + \mathbf{v}) \\
&= M(\mathbf{u} + \mathbf{v}) \bullet M(\mathbf{u} + \mathbf{v}) \\
&= M(\mathbf{u}) \bullet M(\mathbf{u}) + 2M(\mathbf{u}) \bullet M(\mathbf{v}) + M(\mathbf{v}) \bullet M(\mathbf{v}) \\
&= \mathbf{u} \bullet \mathbf{u} + 2M(\mathbf{u}) \bullet M(\mathbf{v}) + \mathbf{v} \bullet \mathbf{v}
\end{aligned}$$

Now cancel the terms $\mathbf{u} \bullet \mathbf{u}$ and $\mathbf{v} \bullet \mathbf{v}$ from both sides and divide by 2.

**2.2.4.4. Theorem.** If M is a motion, then

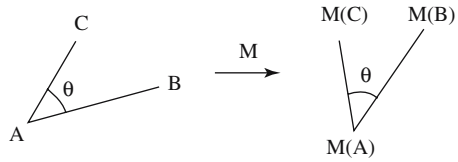
$$M(\mathbf{A})M(\mathbf{B}) \bullet M(\mathbf{A})M(\mathbf{C}) = \mathbf{AB} \bullet \mathbf{AC}$$

for all points $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

**Proof.** By Theorem 2.2.4.2 we can express M in the form $M = TM_0$, where T is a translation and $M_0$ is a motion with $M_0(\mathbf{0}) = \mathbf{0}$. It is easy to check that

$$M(\mathbf{A})M(\mathbf{B}) = M_0(\mathbf{A})M_0(\mathbf{B})$$

**Figure 2.7.** Motions preserve angles.




and

$$M(\mathbf{A})M(\mathbf{C}) = M_0(\mathbf{A})M_0(\mathbf{C}).$$

The theorem now follows from Lemma 2.2.4.3 applied to $M_0$.

**2.2.4.5. Corollary**. Motions preserve angles.

See Figure 2.7. There is a converse to the results proved above.

**2.2.4.6. Theorem.** A map that preserves the length of vectors and the angles between them also preserves distance, that is, it is a motion.

**Proof.** Exercise 2.2.4.1.

## 2.2.5 Some Existence and Uniqueness Results

Let $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_k$ and $\mathbf{P}_1', \mathbf{P}_2', \ldots, \mathbf{P}_k'$, $k \geq 1$, be two sequences of points in the plane. We would like to determine when there is a motion M that sends $\mathbf{P}_i$ to $\mathbf{P}_i'$. Since motions always preserve distances, a minimal requirement is that $|\mathbf{P}_i\mathbf{P}_j| = |\mathbf{P}_i'\mathbf{P}_j'|$ for all i and j. Is this enough though?

**Case 1.** $k = 1$.

There is no problem in this case. For example, the translation $T(\mathbf{Q}) = \mathbf{Q} + \mathbf{P}_1\mathbf{P}_1'$ would do the job. In fact, so would $M = RT$, where R is any rotation about $\mathbf{P}_1'$. In other words, there are an infinite number of distinct motions that send $\mathbf{P}_1$ to $\mathbf{P}_1'$.

**Case 2.** $k = 2$.

Assume, without loss of generality, that $\mathbf{P}_1 \neq \mathbf{P}_2$. Consider the translation T defined in Case 1 that sends $\mathbf{P}_1$ to $\mathbf{P}_1'$. By hypothesis, $|\mathbf{P}_1'T(\mathbf{P}_2)| = |\mathbf{P}_1'\mathbf{P}_2'|$. Let $\alpha$ be the angle between the vectors $\mathbf{P}_1'T(\mathbf{P}_2)$ and $\mathbf{P}_1'\mathbf{P}_2'$ and let R be the rotation about the point $\mathbf{P}_1'$ through the angle $\alpha$. See Figure 2.8. It is easy to show that the motion $M = RT$ does what we want, as does the motion $M' = SRT$, where S is the reflection about the line through $\mathbf{P}_1'$ and $\mathbf{P}_2'$. M and M′ are clearly distinct.

**Case 3.** $k = 3$.

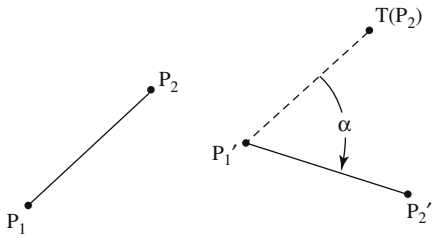



**Figure 2.8.** Moving two points to another two points.

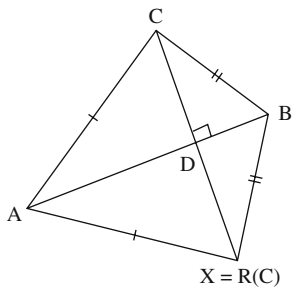



**Figure 2.9.** Proving Lemma 2.2.5.1.

Let M and M′ be the motions defined in Case 2 that send $\mathbf{P}_1$ and $\mathbf{P}_2$ to $\mathbf{P}_1'$ and $\mathbf{P}_2'$, respectively. By hypothesis, $|\mathbf{P}_i'M(\mathbf{P}_3)| = |\mathbf{P}_i'\mathbf{P}_3'|$ for i = 1,2. The next lemma shows that either M or M′ does what we want, namely, either $M(\mathbf{P}_3) = \mathbf{P}_3'$ or $M'(\mathbf{P}_3) = \mathbf{P}_3'$.

**2.2.5.1. Lemma.** Let **A**, **B**, and **C** be three noncollinear points in the plane. The only vectors **X** in the plane that satisfy the two equations $|\mathbf{AX}| = |\mathbf{AC}|$ and $|\mathbf{BX}| = |\mathbf{BC}|$ are $\mathbf{X} = \mathbf{C}$ and $\mathbf{X} = R(\mathbf{C})$, where R is the reflection about the line **L** determined by **A** and **B**.

**Proof.** Assume that $\mathbf{X} \neq \mathbf{C}$.

**Claim.** The midpoint $\mathbf{D} = \frac{1}{2}(\mathbf{C}+\mathbf{X})$ of the segment $[\mathbf{C},\mathbf{X}]$ lies on the line **L**.

See Figure 2.9. Once the claim is proved we are done because the definition of **D** implies that $\mathbf{X} = \mathbf{C} + 2\mathbf{CD}$, which is where the reflection sends the point **D**. Consider the following identities:

$$\begin{aligned}
\mathbf{CD} \bullet \mathbf{AD} &= \left(\frac{1}{2}(\mathbf{C}+\mathbf{X}) - \mathbf{C}\right) \bullet \left(\frac{1}{2}(\mathbf{C}+\mathbf{X}) - \mathbf{A}\right) \\
&= \frac{1}{2}(\mathbf{AX} - \mathbf{AC}) \bullet \frac{1}{2}(\mathbf{AC} + \mathbf{AX}) \\
&= \frac{1}{4}\left(|\mathbf{AX}|^2 - |\mathbf{AC}|^2\right) \\
&= 0
\end{aligned}$$


**Figure 2.10.** Case 4 of the existence theorem.

Similarly, one can show that **CD•BD** = 0. It follows easily from this that the vectors **AD** and **BD** are parallel and that **D** lies on **L**. This proves the claim and the lemma.

**Case 4.** k > 3.

We claim that if the first three points $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ are linearly independent, then the map defined in Case 3 that sends them to $\mathbf{P}_1'$, $\mathbf{P}_2'$, and $\mathbf{P}_3'$, respectively, will already send all the other points $\mathbf{P}_i$, i > 3, to $\mathbf{P}_i'$. Figure 2.10 shows how the argument proceeds. In Figure 2.10(a) we show three circles with centers $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ and radius $\mathbf{P}_1\mathbf{P}_i$, $\mathbf{P}_2\mathbf{P}_i$, and $\mathbf{P}_3\mathbf{P}_i$, respectively. The point $\mathbf{P}_i$ lies on the intersection of these circles. Figure 2.10(b) shows the corresponding circles around the image points. One has to show that $\mathbf{P}_i$ will get sent to the intersection of those circles and that this is the same as the point $\mathbf{P}_i'$.

We have just given a constructive proof of the following theorem.

**2.2.5.2. Theorem.** (The Existence Theorem for Motions) Given points $\mathbf{P}_1$, $\mathbf{P}_2$, . . . , $\mathbf{P}_k$ and $\mathbf{P}_1'$, $\mathbf{P}_2'$, . . . , $\mathbf{P}_k'$ with the property that $|\mathbf{P}_i\mathbf{P}_j| = |\mathbf{P}_i'\mathbf{P}_j'|$ for all i and j, then there is motion M, so that $M(\mathbf{P}_i) = \mathbf{P}_i'$.

**Proof.** See [Gans69] for missing details in the discussion above.

Now that we have answered the question of the existence of certain motions, let us look at the issue of uniqueness more closely?

**2.2.5.3. Theorem.** A motion that has two distinct fixed points fixes every point on the line determined by those points.

**Proof.** Let M be a motion and assume that $M(\mathbf{A},\mathbf{B}) = (\mathbf{A},\mathbf{B})$ for two distinct points **A** and **B**. Let **L** be the line determined by **A** and **B** and let **C** be any point of **L**. If $\mathbf{C} = \mathbf{A} + t\mathbf{AB}$, then Lemma 2.2.2 implies that $M(\mathbf{C}) = M(\mathbf{A}) + tM(\mathbf{A})M(\mathbf{B})$. In other words, $M(\mathbf{C}) = \mathbf{C}$ and the theorem is proved.

**2.2.5.4. Theorem.** Any motion of the plane that leaves fixed three noncollinear points must be the identity.

**Proof.** Let **A**, **B**, and **C** be three noncollinear points and let M be a motion with M(**A**,**B**,**C**) = (**A**,**B**,**C**). Let **P** be any other point in the plane. We would like to show that M(**P**) = **P**. By Theorem 2.2.5.3, M is the identity on the three lines determined by the points **A**, **B**, and **C**. If **P** lies on these lines we are done; otherwise, Lemma 2.2.3 implies that **P** lies on a line through two distinct points that lie on two of these lines. Using Theorem 2.2.5.3 we can again conclude that M fixes **P**.

**2.2.5.5. Corollary.** Two motions of the plane that agree on three noncollinear points must be identical.

**Proof.** Let M and M′ be motions and assume that M(**A**,**B**,**C**) = M′(**A**,**B**,**C**) for three noncollinear points **A**, **B**, and **C**. Consider the motion $T = M^{-1}M'$. Since T(**A**,**B**,**C**) = (**A**,**B**,**C**), Theorem 2.2.5.4 implies that T is the identity, that is, M = M′.

**2.2.5.6. Corollary.** Every motion of the plane is a composite of a translation, a rotation, and/or possibly a reflection.

**Proof.** This follows from the construction in Case 3 above and Corollary 2.2.5.5.

Theorem 2.2.5.3 raises the question whether a motion of the plane that fixes two distinct points is actually the identity map. That is not the case. Reflections, such as the map T(x,y) = (x,–y), can leave all the points of a line fixed but still not be the identity.

**2.2.5.7. Theorem.** A motion M of the plane that fixes two distinct points **A** and **B** is either the identity map or the reflection about the line **L** determined by **A** and **B**.

**Proof**. By Theorem 2.2.5.3, M fixes all the points on the line **L**. Let **C** be any point not on **L**. Lemma 2.2.5.1 shows that **C** gets mapped by M either to itself or to its reflection **C′** about the line **L**. The theorem now follows from the Corollary 2.2.5.5 since we know what M does on three points.

## 2.2.6 Rigid Motions in the Plane

**2.2.6.1. Lemma.** Every rotation R of the plane can be expressed in the form $R = R_0T_1 = T_2R_0$, where $R_0$ is a rotation about the origin and $T_1$ and $T_2$ are translations. Conversely, if $R_0$ is any rotation about the origin through a **nonzero** angle and if T is a translation of the plane, then both $R_0T$ and $TR_0$ are rotations.

**Proof.** Suppose that $R = TR_0T^{-1}$, where $R_0$ is a rotation about the origin and T is a translation. By Theorem 2.2.4.2 we can move the translations to either side of $R_0$, which proves the first part of the lemma. The other part can be proved by showing that certain equations have unique solutions. For example, to show that $TR_0$ is a rotation, one assumes that it is a rotation about some point (a,b) and tries to solve the equations

$$(x - a)\cos\theta - (y - b)\sin\theta + a = x\cos\theta - y\sin\theta + c$$

$$(x-a)\sin\theta+(y-b)\cos\theta+b=x\sin\theta+y\cos\theta+d$$

for a and b. The details are left as an exercise.

**2.2.6.2. Theorem.** The **set** of all translations and rotations of the plane is a subgroup of the group of all motions. The set of rotations by itself is not a group.

**Proof.** To prove the theorem one uses Lemma 2.2.6.1 to show that the composites of translations and rotations about an arbitrary point are again either a translation or a rotation.

**Definition.** A motion of the plane that is a composition of translations and/or rotations is called a *rigid motion* or *displacement*.

Rigid motions are closely related to orientation-preserving maps. We defined that concept in Section 1.6 for linear transformations and we would now like to extend the definition to motions. Motions are not linear transformations, but by Theorem 2.2.4.2 they differ from one by a translation. Intuitively, we would like to say that a motion M of the plane is "orientation preserving" if for every three noncollinear points **A**, **B**, and **C** the ordered pairs of basis vectors (**AB**,**AC**) and (M(**A**)M(**B**),M(**A**)M(**C**)) determine the same orientation of $\mathbf{R}^2$. See Figure 2.11. This definition would be messy to work with and so we take a different approach.

Let M be a motion in $\mathbf{R}^n$. By Theorem 2.2.4.2 we can write M uniquely in the form $M = TM_0$, where T is a translation and $M_0$ is a motion that fixes the origin. Theorem 2.2.4.1 implies that $M_0$ is a linear transformation.

**Definition.** The motion M is said to be *orientation preserving* if $M_0$ is. Otherwise, M is said to be *orientation reversing*.

**2.2.6.3. Theorem.**

(1) A motion M is orientation preserving if and only if $M^{-1}$ is orientation preserving.
(2) The composition $MM'$ of two motions M and $M'$ is orientation preserving if and only if either both are orientation preserving or both are orientation reversing.
(3) The composition $M_1M_2 \ldots M_k$ of motions $M_i$ is orientation preserving if and only if the number of orientation-reversing motions $M_i$ is even.

**Proof.** The proof is left as an exercise. It makes heavy use of Theorem 2.2.4.2 to switch translations from one side of a motion that fixes the origin to the other.

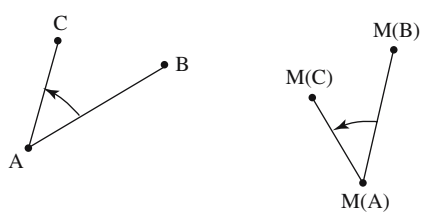



**Figure 2.11.** An orientation-preserving motion.

**2.2.6.4. Theorem.**

(1) Translations and rotations of the plane are orientation-preserving motions.
(2) Reflections are orientation-reversing motions.

**Proof.** The fact that translations are orientation-preserving motions follows immediately from the definition since the identity map is certainly orientation preserving. To prove that rotations are orientation preserving, it suffices to show, by Theorem 2.2.6.3, that any rotation R about the origin is orientation preserving since an arbitrary rotation is a composition of translations and a rotation about the origin. The fact that such an R is orientation preserving follows from Theorems 1.6.6 and 2.2.2.1 and the fact that the matrix for the linear transformation R has determinant +1. This proves (1).

To prove (2) note that the reflection $S_x$ about the x-axis is a linear transformation with equation (2.12) that clearly has determinant –1 and hence is orientation reversing. Next, Theorem 2.2.3.6 showed that an arbitrary reflection can be written in the form $T^{-1}R^{-1}S_xRT$, where T is a translation and R is a rotation about the origin. Property (2) now follows from (1) and Theorem 2.2.6.3.

We can also justify Theorem 2.2.6.4(2) geometrically based on the intuitive idea mentioned earlier that a motion of the plane is orientation reversing if for some three noncollinear points **A**, **B**, and **C** the ordered pairs of basis vectors (**AB**,**AC**) and (M(**A**)M(**B**),M(**A**)M(**C**)) determine opposite orientations for $\mathbf{R}^2$. To see this we shall use the same notation as in the definition of a reflection in Section 2.2.3. If **P** is a point not on **L**, then clearly **AQ** and **QP** form a basis for $\mathbf{R}^2$ and

$$\begin{aligned} T(\mathbf{A})T(\mathbf{Q}) &= \mathbf{AQ} = 1 \cdot \mathbf{AQ} + 0 \cdot \mathbf{QP} \\ T(\mathbf{Q})T(\mathbf{P}) &= \mathbf{QP'} = 0 \cdot \mathbf{AQ} + (-1) \cdot \mathbf{QP}. \end{aligned}$$

The determinant of the matrix of coefficients that relates the original basis to the transformed one is –1. This means that the two bases are in opposite orientation classes.

**2.2.6.5. Theorem.** A motion of the plane is orientation preserving if and only if it is a rigid motion.

**Proof.** Exercise.

Although it takes three points to specify a general motion of the plane, two points suffice in the special case of rigid motions.

**2.2.6.6. Theorem.** If M is a rigid motion of the plane and if M fixes two distinct points, then M is the identity.

**Proof.** This theorem is an immediate consequence of Theorems 2.2.5.7 and 2.2.6.4.

**2.2.6.7. Corollary.** Two rigid motions of the plane that agree on two distinct points must be identical.

**Proof.** The proof of this corollary is similar to the proof of Corollary 2.2.5.5.

**2.2.6.8. Corollary.** If two orientation-reversing motions of the plane agree on two distinct points, then they must be identical.

**Proof.** If M and M′ are the two orientation-reversing motions, then $M' M^{-1}$ is a rigid motion that fixes two distinct points and hence is the identity map. It follows that M = M′.

**2.2.6.9. Theorem.** A rigid motion of the plane that has a fixed point **p** is a rotation about **p**.

**Proof.** Exercise.

## 2.2.7 Summary for Motions in the Plane

We have defined motions and have shown that a motion of the plane is completely specified by what it does to three noncollinear points and that it can be described in terms of three very simple motions, namely, translations, rotations, and reflections. To understand such motions it suffices to have a good understanding of these three primitive types.

Planar motions are either orientation preserving or orientation reversing with rigid motions being the orientation-preserving ones. Reflections are orientation reversing. Another way to describe a planar motion is as a rigid motion or the composition of a rigid motion and a single reflection. In fact, we may assume that the reflection, if it is needed, is just the reflection about the x-axis.

Combining various facts we know, it is now very easy to describe the equation of an arbitrary motion of the plane.

**2.2.7.1. Theorem.** Every motion M of the plane is defined by equations of the form

$$\begin{aligned} x' &= ax + by + c \\ y' &= \pm(-bx + ay) + d, \end{aligned} \tag{2.19}$$

where $a^2 + b^2 = 1$. Conversely, every such pair of equations defines a motion.

**Proof.** Let M(**0**) = (c,d) and define a translation T by T(**P**) = **P** + (c,d). Let M′ = $T^{-1}$ M. Then M = TM′ and M′ fixes the origin.

**Case 1.** M is orientation preserving.

In this case M′ is orientation preserving and must be a rotation about the origin through some angle θ (Theorem 2.2.6.9). Let $a = \cos\theta$ and $b = -\sin\theta$. Clearly the equation for M has the desired form.

**Case 2.** M is orientation reversing.

Let S be the reflection about the x-axis, that is, $S(x,y) = (x,-y)$. Since M′ is orientation reversing, it follows that the motion $R = SM'$ is orientation preserving, but R also fixes the origin. Therefore, R must be a rotation about the origin through some angle θ. Note that $SR = SSM' = M'$. Define a and b as in Case 1. It is again easy to see that the equation for $M = TM' = TSR$ has the desired form.

This proves the first part of Theorem 2.2.7.1. The second part is Exercise 2.2.7.1. See also the next example.

**2.2.7.2. Example.** Let us show directly, without using Theorem 2.2.7.1, that the transformation M defined by the equations

$$x' = \frac{\sqrt{3}}{2}x + \frac{1}{2}y + 5$$

$$y' = \frac{1}{2}x - \frac{\sqrt{3}}{2}y + 7$$

is a motion.

**Solution.** Define a translation T by $T(x,y) = (x,y) + (5,7)$. Let R be the rotation about the origin through the angle $-\pi/6$ and let S be the reflection about the x-axis. It is easy to see that $M = TSR$ and hence is a motion since it is a composite of motions.

Theorem 2.2.7.1 shows that motions can be represented by five real numbers (the a, b, c, d, and ±1 depending on the sign). Rigid motions can be represented by four real numbers. Chapter 20 in [AgoM05] describes a very compact way to represent motions in terms of quaternions. The fact that a motion is defined by five numbers leads to another way to solve for a motion when it is given in terms of some points and their images. One simply solves the equations in Theorem 2.2.7.1 for the unknown coefficients. Solving for five unknowns turns out to be not as complicated as it may sound in this case.

Next, we would like to give a more complete geometric characterization of motions than that given in Corollary 2.2.5.6.

**2.2.7.3. Lemma.** Every orientation-reversing motion M that fixes the origin is a reflection about a line through the origin.

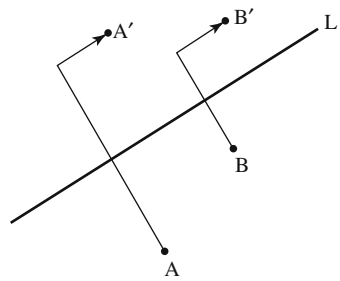
**Proof.** Let **p** be a nonzero point. If M fixes **p**, then Theorem 2.2.5.7 implies that M is the reflection about the line through the origin and **p** and we are done. Assume therefore that $\mathbf{p}' = M(\mathbf{p}) \neq \mathbf{p}$. Let **q** be the midpoint of the segment $[\mathbf{p},\mathbf{p}']$ and let S be the reflection about the line through the origin and **q**. Clearly, S and M agree on **0** and **p**. By Corollary 2.2.6.8 they are the same map.

**Definition.** A *glide reflection* is the composite of a reflection about a line **L** followed by a translation with **nonzero** translation vector parallel to **L**. (See Figure 2.12.)

**2.2.7.4. Theorem.** Every orientation-reversing motion M is a reflection or a glide reflection.

**Proof.** If M fixes the origin, then the theorem is true by Lemma 2.2.7.3. Assume therefore that $M(\mathbf{0}) = \mathbf{p}$ is distinct from the origin and let T be the translation that

**Figure 2.12.** A glide reflection.




sends the origin to **p.** Let $M' = T^{-1}M$. Then $M'$ is an orientation-reversing motion that fixes the origin and hence a reflection by Lemma 2.2.7.3. Since $M = TM'$, M is a glide reflection and we are done.

**2.2.7.5. Theorem.** Every motion M is either a translation, rotation, reflection, or glide reflection.

**Proof.** If M is a rigid motion, then M is a translation or rotation by Theorem 2.2.6.2. If M is not a rigid motion, that is, if it is orientation reversing, then M is a reflection or glide reflection by Theorem 2.2.7.4.

One final word about why the term "congruent transformation" is sometimes used instead of "motion." The reader may recall the notion of "congruent figures" from his/her Euclidean geometry course in high school, which most likely was never given a really precise definition. Well, we can do so now.

**Definition.** Two figures are said to be *congruent* if there is a motion that carries one into the other.

## 2.2.8 Frames in the Plane

Before leaving the subject of motions in the plane we want to discuss another approach to defining them – one that will be especially powerful in higher dimensions.

**Definition.** A *frame* in $\mathbf{R}^2$ is a tuple $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{p})$, where **p** is a point and $\mathbf{u}_1$ and $\mathbf{u}_2$ define an orthonormal basis of $\mathbf{R}^2$. If the ordered basis $(\mathbf{u}_1,\mathbf{u}_2)$ induces the standard orientation, then we shall call the frame an *oriented frame*. The lines determined by **p** and the direction vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are called the *x*-, respectively, *y-axis* of the frame F. The point **p** is called the *origin* of the frame F. $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{0})$ is called the *standard frame* of $\mathbf{R}^2$. To simplify the notation, we sometimes use $(\mathbf{u}_1,\mathbf{u}_2)$ to denote the frame $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{0})$.

Frames can be thought of as defining a new coordinate system. See Figure 2.13. They can also be associated to a transformation in a natural way. If $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{p})$ is a frame and if $\mathbf{u}_i = (u_{i1},u_{i2})$ and $\mathbf{p} = (m,n)$, then define a map $T_F$ by the equations

$$\begin{aligned} x' &= u_{11}x + u_{21}y + m \\ y' &= u_{12}x + u_{22}y + n. \end{aligned} \tag{2.20}$$

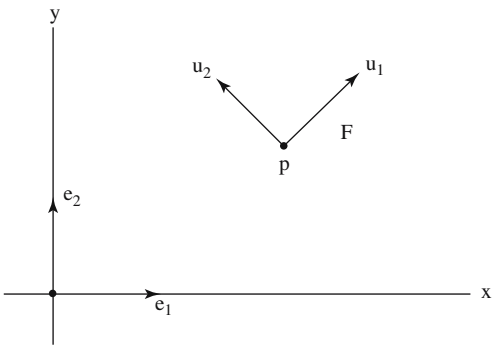



**Figure 2.13.** Frames in the plane.

In matrix form, $T_F$ is the map

$$T_F(x,y) = (x,y)\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} + \mathbf{p}. \tag{2.21}$$

**Claim 1.** $T_F$ is a motion.

**Proof.** Since $(\mathbf{u}_1,\mathbf{u}_2)$ is an orthonormal basis, we have that

$$u_{11}^2 + u_{12}^2 = 1 = u_{21}^2 + u_{22}^2 \quad \text{and} \quad u_{11}u_{21} + u_{12}u_{22} = 0.$$

If follows easily from this that the equations (2.20) have the form of the equations in Theorem 2.2.7.1, proving the claim.

If we think of a frame as defining a new coordinate system, then we can coordinatize the points in the plane with respect to it.

**Definition.** The coordinates of a point with respect to a frame are called the *frame coordinates*. The frame coordinates with respect to the standard frame are called *world coordinates*.

Since $T_F$ maps the origin (0,0) to $\mathbf{p}$, (1,0) to $\mathbf{p} + \mathbf{u}_1$, and (0,1) to $\mathbf{p} + \mathbf{u}_2$, we can think of $T_F$ as mapping frame coordinates to world coordinates.

There is a converse to Claim 1. Let M be a motion defined by the equations

$$x' = ax + by + m$$
$$y' = cx + dy + n.$$

Let $\mathbf{u}_1 = (a,c)$, $\mathbf{u}_2 = (b,d)$, and $\mathbf{p} = (m,n)$.

**Claim 2.** $(\mathbf{u}_1,\mathbf{u}_2)$ is an orthonormal basis and $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{p})$ is a frame.

**Proof.** This also follows from Theorem 2.2.7.1 since by that theorem c = –kb and d = ka for some k = ±1 and $a^2 + b^2 = 1$.

Claims 1 and 2 can be summarized by saying that there is a one-to-one correspondence between frames and motions. The special case where $\mathbf{p} = \mathbf{0}$ shows that there is a one-to-one correspondence between orthonormal bases and motions that fix $\mathbf{0}$.

**2.2.8.1. Example.** Consider the rotation R about the origin defined by

$$x' = \frac{3}{5}x - \frac{4}{5}y$$

$$y' = \frac{4}{5}x + \frac{3}{5}y.$$

The vectors $\mathbf{u}_1 = \left(\frac{3}{5}, -\frac{4}{5}\right)$ and $\mathbf{u}_2 = \left(\frac{4}{5}, \frac{3}{5}\right)$ clearly form an orthonormal basis.

**Definition.** The motion $T_F$, usually simply denoted by F, is called *the motion defined by F*.

Using "F" to denote both the frame F and the motion $T_F$ should not cause any confusion since it will always be clear from the context as to whether we are talking about the frame or the map.

The observations above lead to a simple way to get the inverse of a motion. Consider equations (2.20) again. Let R be the motion

$$R(x, y) = (x, y)\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}$$

and T, the translation

$$T(\mathbf{q}) = \mathbf{q} + (m, n).$$

(Note that R is actually a rotation if the frame is oriented.) Then, as maps, F = TR and $F^{-1} = R^{-1}T^{-1}$. But it is easy to check that

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}\left(\mathbf{u}_1^T \ \mathbf{u}_2^T\right) = (\mathbf{u}_i \bullet \mathbf{u}_j) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which shows that the inverses of the matrices

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \quad \text{and} \quad \left(\mathbf{u}_1^T \ \mathbf{u}_2^T\right)$$

are just their transposes. Considering Example 2.2.8.1 again, note that the transpose of

$$\begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}$$

is its inverse. Therefore, $F^{-1}$ is the map defined by

$$F^{-1}(x, y) = ((x, y) - \mathbf{p})\left(\mathbf{u}_1^T\ \mathbf{u}_2^T\right), \tag{2.22}$$

or, in terms of equations,

$$\begin{aligned} x' &= u_{11}(x - m) + u_{12}(y - n) \\ y' &= u_{21}(x - m) + u_{22}(y - n). \end{aligned} \tag{2.23}$$

Equations (2.20)–(2.23) are fundamental and worth remembering. They summarize the main relationship between frames and motions.

We finish this discussion of frames with several examples.

**2.2.8.2. Example.** To find the equations for the rotation about the origin which rotates the point $\mathbf{A} = (2,0)$ into $\mathbf{B} = (1,\sqrt{3})$.

**Solution.** All we have to do is to normalize $\mathbf{B}$ to

$$\mathbf{u}_1 = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

and combine this vector with the orthogonal vector

$$\mathbf{u}_2 = \left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$$

(chosen so that the pair induces the standard orientation) to get the frame $F = (\mathbf{u}_1, \mathbf{u}_2)$. This frame defines the desired rotation.

**2.2.8.3. Example.** To find a motion M that sends the origin to the point $\mathbf{A} = (3,0)$ and the directed x-axis to the directed line $\mathbf{L}_1$ shown in Figure 2.14.

**Solution.** Define a frame $F = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{A})$ by

$$\mathbf{u}_1 = \frac{\mathbf{AB}}{|\mathbf{AB}|} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right),$$

$$\mathbf{u}_2 = \left(-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right).$$

Then M = F does the job. In fact, M is a rigid motion.

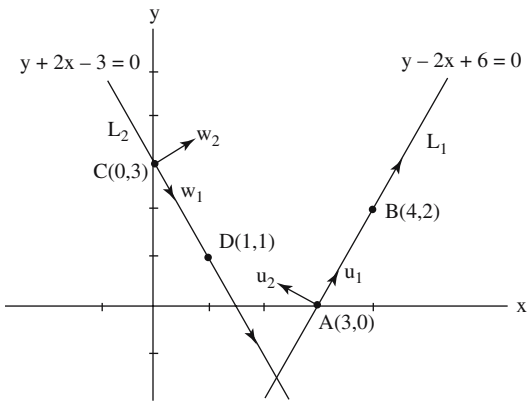



**Figure 2.14.** Using frames to move lines.

Example 2.2.8.3 easily generalizes to finding a motion that maps the x-axis and the origin to any other line and point. By taking the inverse of this map we can map an arbitrary line to the x-axis. We can go a step further:

**2.2.8.4. Example.** Again consider Figure 2.14. We find a motion M that sends the point $\mathbf{A} = (3,0)$ to the point $\mathbf{C} = (0,3)$ and the directed line $\mathbf{L}_1$ to the directed line $\mathbf{L}_2$.

**Solution.** Following the approach used in Example 2.2.8.3 we can map the x-axis to the line $\mathbf{L}_2$ using the map G, where G is the frame $(\mathbf{w}_1,\mathbf{w}_2,\mathbf{C})$ and

$$\mathbf{w}_1 = \frac{\mathbf{CD}}{|\mathbf{CD}|} = \left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right),$$

$$\mathbf{w}_2 = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right).$$

If F is the frame defined in Example 2.2.8.3, then $M = GF^{-1}$ is a rigid motion that will do what we want. In terms of equations we have

$$F^{-1}: \quad x' = \frac{1}{\sqrt{5}}(x-3) + \frac{2}{\sqrt{5}}y$$

$$y' = -\frac{2}{\sqrt{5}}(x-3) + \frac{1}{\sqrt{5}}y,$$

$$G: \quad x' = \frac{1}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y$$

$$y' = -\frac{2}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y + 3, \quad \text{and}$$

$$M: \quad x' = -\frac{3}{5}x + \frac{4}{5}y + \frac{9}{5}$$

$$y' = -\frac{4}{5}x - \frac{3}{5}y + \frac{27}{5}.$$

Looking over the solution for Example 2.2.8.4, a reader might wonder if there was a special reason for choosing the particular orthonormal bases $(\mathbf{u}_1,\mathbf{u}_2)$ and $(\mathbf{w}_1,\mathbf{w}_2)$. Not really. The important observation is that no matter what the choice, the three points $\mathbf{A}$, $\mathbf{A}+\mathbf{u}_1$, and $\mathbf{A}+\mathbf{u}_2$ will get sent to the points $\mathbf{C}$, $\mathbf{C}+\mathbf{w}_1$, and $\mathbf{C}+\mathbf{w}_2$. The line determined by the points $\mathbf{A}$ and $\mathbf{A}+\mathbf{u}_1$ will get sent to the line determined by $\mathbf{C}$ and $\mathbf{C}+\mathbf{w}_1$. We could have replaced the bases by $(\mathbf{u}_1, \pm\mathbf{u}_2)$ and $(\mathbf{w}_1, \pm\mathbf{w}_2)$ or $(-\mathbf{u}_1, \pm\mathbf{u}_2)$ and $(-\mathbf{w}_1, \pm\mathbf{w}_2)$ and we would have gotten an answer to our problem. On the other hand, if we want to get a rigid motion then things are not quite so arbitrary. We still have choices, but the bases must induce the same orientation of the plane. In particular, we would not be able to pick $(\mathbf{u}_1,\mathbf{u}_2)$ and $(\mathbf{w}_1,-\mathbf{w}_2)$, for example.

It is easy to see from Example 2.2.8.4 that frames can easily be used to solve the general problem of mapping one directed line and point to another directed line and point. The user should compare this approach to how the problem would be solved without frames. The underlying mathematics is really the same. The orthonormal bases have the cosines of angles that are used for the rotation contained in them implicitly. Recall that the components of a unit vector are just direction cosines. Nevertheless, with frames one simply has to build orthonormal bases and this is easier than messing with angles directly.

Finally, one can also use frames to define motions that send three points to another three points. For example, suppose that we want to define a motion M that sends points $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ to points $\mathbf{A}'$, $\mathbf{B}'$, and $\mathbf{C}'$, respectively. See Figure 2.15. Let $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{A})$ be the frame obtained from the normalized $\mathbf{AB}$ and the orthogonal complement of $\mathbf{AC}$ with respect to $\mathbf{AB}$. Let $F' = (\mathbf{u}_1',\mathbf{u}_2',\mathbf{A}')$ be the frame obtained in a similar way from $\mathbf{A}'$, $\mathbf{B}'$, and $\mathbf{C}'$. Then $M = F'F^{-1}$.

**2.2.8.5. Example.** To find the motion M that sends $\mathbf{A}(-2,1)$, $\mathbf{B}(0,2)$, $\mathbf{C}(-2,4)$ to $\mathbf{A}'(4,0)$, $\mathbf{B}'(6,-1)$, $\mathbf{C}'(4,-3)$, respectively.

**Solution.** See Figure 2.16. The first task is to define the frames $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{A})$ and $F' = (\mathbf{u}_1',\mathbf{u}_2',\mathbf{A}')$ so that $M = F'F^{-1}$. To get the orthonormal bases we apply the Gram-Schmidt algorithm to the bases $(\mathbf{AB},\mathbf{AC})$ and $(\mathbf{A'B'},\mathbf{A'C'})$. We leave this as an exercise for the reader. One gets

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}(2,1), \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}}(-1,2)$$

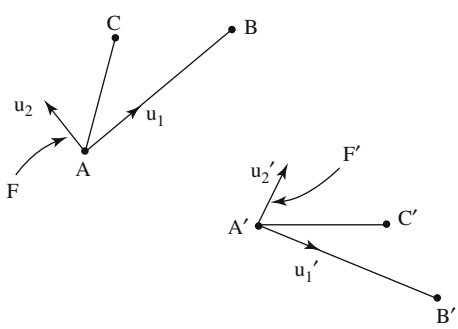
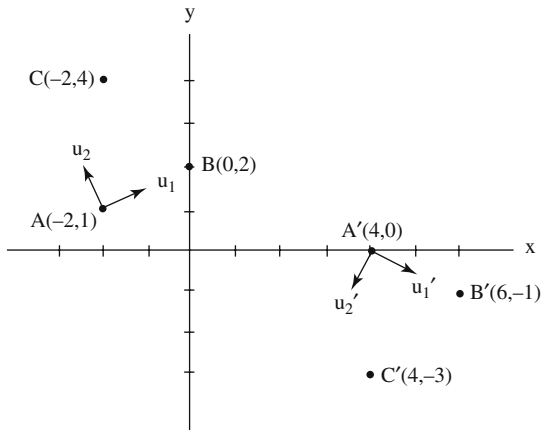



**Figure 2.15.** Using frames to define motions.

**Figure 2.16.** Example 2.2.8.5.




and

$$\mathbf{u}_1' = \frac{1}{\sqrt{5}}(2,1), \quad \mathbf{u}_2' = \frac{1}{\sqrt{5}}(-1,2).$$

Equations (2.22) and (2.21) imply that

$$F^{-1}(x,y) = (x+2, y-1)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix},$$

$$F'(x,y) = (x,y)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix} + A',$$

and

$$(x',y') = M(x,y) = (x+2, y-1)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix} + (4,0).$$

This leads to the following equations for M:

$$x' = x + 6$$
$$y' = -y + 1$$

It is easy to check that this motion sends the points **A**, **B**, **C** to **A′**, **B′**, **C′**.

The astute reader may have noticed just by looking at Figure 2.16 that there are easier ways to solve Problem 2.2.8.5. For example, the motion M can also be obtained by translating **A** to **A′** and then reflecting about the x-axis. However, to emphasize a point made earlier, using frames is a **systematic** approach that can be programmed on a computer. Computers cannot "look."

## 2.3 Similarities

**Definition.** A map $S:\mathbf{R}^n \to \mathbf{R}^n$ is called a *similarity transformation*, or simply a *similarity*, if

$$|S(\mathbf{p})S(\mathbf{q})| = r|\mathbf{pq}|$$

for all $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$ and some fixed **positive** constant r.

Clearly, motions are similarities, because they correspond to the case where r is 1 in the definition. On the other hand, the map $S(\mathbf{p}) = 2\mathbf{p}$ is a similarity but not a motion. In fact, S an example of a simple but important class of similarities.

**Definition.** A map $R:\mathbf{R}^n \to \mathbf{R}^n$ of the form $R(\mathbf{p}) = r\mathbf{p}$, $r > 0$, is called a *radial transformation*.

**2.3.1. Theorem.** Radial transformations are similarities.

**Proof.** Exercise.

The next theorem shows that similarities are not much more complicated than motions.

**2.3.2. Theorem.** If S is a similarity, then S = MR, where M is a motion and R is a radial transformation. Conversely, any map of the form MR, where M is a motion and R is a radial transformation, is a similarity.

**Proof.** This is easy because if we use the notation in the definitions for a similarity and a radial transformation, then $R^{-1}S$ is a motion M.

**2.3.3. Corollary.** Every similarity in the plane can be expressed by equations of the form

$$x' = ax + by + m$$

$$y' = \pm(-bx + ay) + n, \tag{2.24}$$

where $(a,b) \neq (0,0)$. (The r in the definition of a similarity is $\sqrt{a^2 + b^2}$ in this case.) Conversely, every map defined by such equations is a similarity.

**2.3.4. Theorem.** Similarity transformations

(1) preserve the betweenness relation,
(2) preserve collinearity and noncollinearity,
(3) map lines onto lines, and
(4) are one-to-one transformations of $\mathbf{R}^n$ onto itself.

**Proof.** This theorem follows from Theorem 2.3.2 and some obvious facts about radial transformations. In the planar case, it can also be proved directly like it was done in the case of motions.

**2.3.5. Theorem.** The similarity transformations form a group that contains the group of motions as a subgroup.

**Proof.** Obvious.

**2.3.6. Theorem.** A similarity transformation in the plane is completely specified by its action on three noncollinear points.

**Proof.** Use Theorem 2.3.2.

**2.3.7. Theorem.** Similarity transformations in the plane preserve angles.

**Proof.** By Theorem 2.3.2, since motions preserve angles, it suffices to show that radial transformations preserve angles, which is easy.

## 2.4 Affine Transformations

**Definition.** A one-to-one and onto mapping $T:\mathbf{R}^n \to \mathbf{R}^n$ that maps lines onto lines is called an *affine transformation*.

Actually, one can characterize affine transformations in a slightly stronger fashion.

**2.4.1. Theorem.** Any one-to-one and onto map of $\mathbf{R}^n$ onto itself that preserves collinearity is an affine transformation.

**Proof.** The only thing that needs to be shown is that lines get mapped **onto** lines. This is shown in a way similar to what was done in the proof of Lemma 2.2.4 and left as an exercise.

**2.4.2. Theorem.** The set of affine transformations in $\mathbf{R}^n$ forms a group that contains the similarities as a subgroup.

**Proof.** Exercise.

Affine transformations, like motions and similarities, have a simple analytic description. Before we get to the main result for these maps in the plane, we analyze transformations with equations of the form

$$\begin{aligned} x' &= ax + by + m \\ y' &= cx + dy + n \end{aligned} \tag{2.25a}$$

where

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0. \tag{2.25b}$$

**2.4.3. Theorem.** The set of transformations defined by equations (2.25) form a group under composition.

**Proof.** This is straightforward. The main observation is that since the determinant in (2.25b) is nonzero, the transformations have inverses. It is also easy to show that the inverses are defined by equations of the same form.

The transformations defined by equations (2.25) clearly include the motions and similarities. It is worth noting that they are simply the composition of a linear transformation of the plane followed by a translation. There are two other interesting special cases.

**Definition.** The linear transformation of the plane defined by the equations

$$\begin{aligned} x' &= ax \\ y' &= \quad\ \ dy, \quad ad \neq 0, \end{aligned} \tag{2.26}$$

is called a (*local*) *scaling transformation*. It is a *global scaling transformation* if $a = d$.

Note that the scaling transformation defined by equations (2.26) is orientation reversing if $ad < 0$. It will be a similarity if $a = d > 0$. It is easy to check that the inverse of the scaling transformation above is the scaling transformation

$$\begin{aligned} x' &= (1/a)x \\ y' &= \qquad (1/d)y. \end{aligned}$$

**Definition.** A linear transformation of the plane defined by equations

$$\begin{aligned} x' &= \ \ x \\ y' &= cx + y \end{aligned} \tag{2.27a}$$

is called a *shear in the x-direction*. A linear transformation defined by equations

$$\begin{aligned} x' &= x + by \\ y' &= \quad\ \ y \end{aligned} \tag{2.27a}$$

is called a *shear in the y-direction*.

It is easy to show that the inverse of a shear is a shear. See Exercise 2.4.1.

**2.4.4. Theorem.** Every transformation of the plane defined by equations (2.25) is a composition of translations, rotations, shears, and/or scaling transformations. Conversely, every composition of such maps can be described by equations of the form (2.25).

**Proof.** Let M be defined by equations (2.25) and set $M_0 = TM$, where T is the translation with translation vector $(-m,-n)$. Then $M_0(\mathbf{0}) = \mathbf{0}$ and $M_0$ is a nonsingular linear transformation. Let $r = |M_0(\mathbf{e}_1)|$, let R be the rotation about the origin that rotates the unit vector $(1/r)M_0(\mathbf{e}_1)$ into $\mathbf{e}_1$, and let $M_1 = RM_0$. It follows that $M_1$ is defined by equations

$$M_1(\mathbf{e}_1) = \mathbf{v}_1 = r\mathbf{e}_1$$
$$M_1(\mathbf{e}_2) = \mathbf{v}_2 = s\mathbf{e}_1 + t\mathbf{e}_2.$$

Define a linear transformation S by

$$S(\mathbf{v}_1) = \mathbf{v}_1$$
$$S(\mathbf{v}_2) = (\mathbf{v}_2 \bullet \mathbf{e}_2)\mathbf{e}_2 = \mathbf{w}_2 = t\mathbf{e}_2.$$

Since a linear transformation is completely defined once it is defined on a basis, S is well defined. In fact, it is easy to show that S is a shear in the x-direction defined by equations

$$x' = x$$
$$y' = -\frac{s}{t}x + y.$$

Figure 2.17 shows the effect of the maps R and S. The map $M_2 = SM_1$ is now the scaling transformation defined by

$$x' = rx$$
$$y' = sy.$$

To summarize, $M = T^{-1}R^{-1}S^{-1}M_2$ and the first part of the theorem is proved. Since the converse of the theorem is obvious, Theorem 2.4.4 is proved.

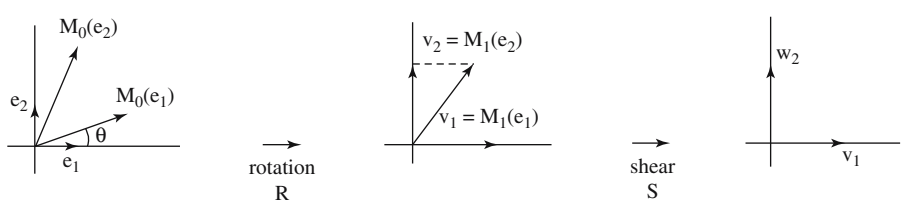



**Figure 2.17.** The rotation and shear in the proof of Theorem 2.4.4.

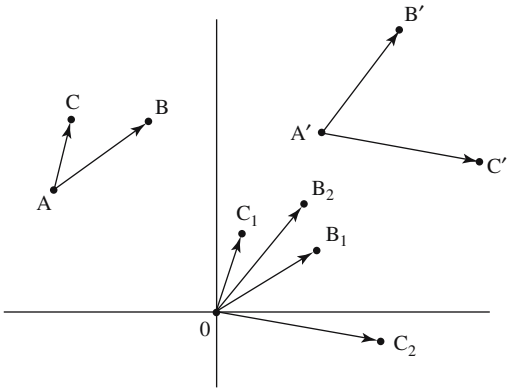



**Figure 2.18.** Proving Theorem 2.4.5.

**2.4.5. Theorem.** Any three noncollinear points in the plane can be mapped into any other three noncollinear points by a unique transformation M with equations (2.25).

**Proof.** Let $(\mathbf{A},\mathbf{B},\mathbf{C})$ and $(\mathbf{A}',\mathbf{B}',\mathbf{C}')$ be two triples of noncollinear points. Let $T_1$ and $T_2$ be the translations that send $\mathbf{A}$ and $\mathbf{A}'$, respectively, to the origin $\mathbf{0}$. Let $T_1(\mathbf{A},\mathbf{B},\mathbf{C}) = (\mathbf{0},\mathbf{B}_1,\mathbf{C}_1)$ and $T_2(\mathbf{A}',\mathbf{B}',\mathbf{C}') = (\mathbf{0},\mathbf{B}_2,\mathbf{C}_2)$. See Figure 2.18. Since the vectors $\mathbf{B}_1$ and $\mathbf{C}_1$ and the vectors $\mathbf{B}_2$ and $\mathbf{C}_2$ are bases for $\mathbf{R}^2$, there is a linear transformation $M_0$ with $M_0(\mathbf{B}_1,\mathbf{C}_1) = (\mathbf{B}_2,\mathbf{C}_2)$. The map $M = T_2^{-1}M_0T_1$ does what we want.

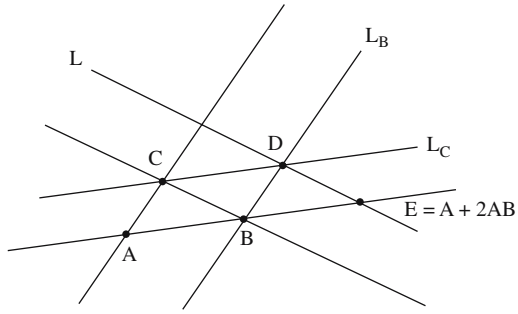
To prove the uniqueness of M, suppose there is another such map M′, then $M^{-1}M'(\mathbf{A},\mathbf{B},\mathbf{C}) = (\mathbf{A},\mathbf{B},\mathbf{C})$. As usual, it therefore suffices to show that any map T defined by equations (2.25) that fixes three noncollinear points $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ is the identity map. There are many ways to prove this. For example, we may assume that $\mathbf{A} = \mathbf{0}$, so that T is linear transformation. Then $\mathbf{B}$ and $\mathbf{C}$ are a basis for the plane and since T is defined by what it does on a basis, it must be the identity everywhere.

We return to affine transformations.

**2.4.6. Theorem.** An affine transformation of the plane that fixes three noncollinear points is the identity map.

**Sketch of proof.** Suppose that the affine transformation T fixes the noncollinear points $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. The property of T we shall use over and over again is that if $\mathbf{P}$ and $\mathbf{Q}$ are distinct points, then T maps the line through $\mathbf{P}$ and $\mathbf{Q}$ into the line through $T(\mathbf{P})$ and $T(\mathbf{Q})$. Let $\mathbf{L_B}$ be the line through $\mathbf{B}$ that is parallel to the line through $\mathbf{A}$ and $\mathbf{C}$. Let $\mathbf{L_C}$ be the line through $\mathbf{C}$ that is parallel to the line through $\mathbf{A}$ and $\mathbf{B}$. See Figure 2.19. Then $T(\mathbf{L_B}) \subseteq \mathbf{L_B}$ and $T(\mathbf{L_C}) \subseteq \mathbf{L_C}$ because parallelism is preserved. It follows that if $\mathbf{D}$ is the intersection of the lines $\mathbf{L_B}$ and $\mathbf{L_C}$, then $T(\mathbf{D}) = \mathbf{D}$. Next, let $\mathbf{L}$ be the line through $\mathbf{D}$ that is parallel to the line through $\mathbf{B}$ and $\mathbf{C}$ and let $\mathbf{E}$ be the intersection of $\mathbf{L}$ and the line through $\mathbf{A}$ and $\mathbf{B}$. Clearly, $\mathbf{E} = \mathbf{A} + 2\mathbf{AB}$. By an argument similar to the one that showed that T fixes $\mathbf{D}$, we must have $T(\mathbf{E}) = \mathbf{E}$. It is easy to continue this type of argument to show that T fixes all points in the form $\mathbf{A} + m\mathbf{AB} + n\mathbf{AC}$, $m, n \in \mathbf{Z}$. From here it is only a small step to show that T fixes all points in the

**Figure 2.19.** Proving Theorem 2.4.6.




form **A** + r**AB** + s**AC** for all rational numbers r and s. These points are a dense set of points in the plane. The final step handles the points where r or s are irrational. See [Gans69].

**2.4.7. Corollary.** An affine transformation of the plane is completely determined by what it does to three noncollinear points.

**Proof.** Showing that the corollary follows from Theorem 2.4.6 uses an, by now standard, argument that is left as an exercise for the reader.

We are ready to state and prove a fundamental theorem about affine maps.

**2.4.8. Theorem.** Every affine transformation of the plane can be described uniquely by equations of the form (2.25). The determinant in (2.25b) is called the *determinant of the affine transformation*. Conversely, every such pair of equations defines an affine transformation.

**Proof.** We start with the converse. A transformation T defined by equations (2.25) has an inverse that is again defined by linear equations of the same form. Let f(x,y) = 0 be the equation of a line **L**. Then the set **L′** = T(**L**) is defined by the equation $f(T^{-1}(x,y)) = 0$. This shows that **L′** is a line and that T is an affine map.

Next, let T be an affine map and choose three noncollinear points. By Theorem 2.4.5 there is a map M defined by equations (2.25) that agrees with T on those points. Since we just showed that T is an affine map, we have two affine maps that act the same on three noncollinear points. By Corollary 2.4.7, T = M and the theorem is proved.

Because of Theorem 2.4.8 everything proved for the maps defined by equations (2.25) holds for affine maps. We restate these properties to emphasize their validity for affine maps.

(1) Every affine map in the plane is a composition of translations, rotations, shears, and/or scaling transformations. Conversely, every composition of such maps is an affine map.
(2) There is a unique affine transformation in the plane that maps three noncollinear points into any other three noncollinear points.

**2.4.9. Theorem.** The only affine transformations of the plane that preserve angles are similarities.

**Sketch of proof.** Let T be an affine transformation that preserves angles. Choose noncollinear points **A**, **B**, and **C**. If T(**A**,**B**,**C**) = (**A′**,**B′**,**C′**), then one can show that

$$|\mathbf{A'B'}| = r|\mathbf{AB}|, \quad |\mathbf{B'C'}| = r|\mathbf{BC}|, \quad \text{and} \quad |\mathbf{A'C'}| = r|\mathbf{AC}|$$

for some r > 0. Let U be the radial transformation U(**p**) = (1/r)**p** and let (**A″**,**B″**,**C″**) = (UT)(**A**,**B**,**C**). There is a unique motion M such that (**A″**,**B″**,**C″**) = M(**A**,**B**,**C**). Now S = $U^{-1}M$ is a similarity that agrees with T on **A**, **B**, and **C**. By Corollary 2.4.7, T and S must be the same transformations.

**Definition.** The *ratio of division* of three **distinct** points **A**, **B**, and **P** on an oriented line **L** in $\mathbf{R}^n$, denoted by (**AB**,**P**), is defined by

$$(\mathbf{AB},\mathbf{P}) = \frac{\|\mathbf{AP}\|}{\|\mathbf{PB}\|}.$$

(||**AP**|| and ||**PB**|| are the signed distances on the oriented line **L**.)

**2.4.10. Proposition.** Let **A**, **B**, and **P** be distinct points on an oriented line **L**. If **P** = **A** + t**AB** = (1 − t)**A** + t**B**, then

$$(\mathbf{AB},\mathbf{P}) = \frac{t}{1-t}.$$

In particular, (**AB**,**P**) is independent of the orientation of **L**.

**Proof.** See Figure 2.20. The proof is a straightforward consequence of the fact that **AP** = t**AB** and **PB** = (1 − t)**AB**.

Using Proposition 2.4.10 it is easy to show that the ratio of division (**AB**,**P**) is positive if **P** belongs to the segment [**A**,**B**] and negative otherwise.

**2.4.11. Proposition.** Let T be an affine transformation of the plane. If **A**,**B** ∈ $\mathbf{R}^2$, then

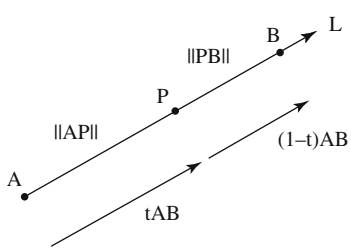



**Figure 2.20.** The ratio of division.

$$T((1-t)\mathbf{A}+t\mathbf{B}) = (1-t)T(\mathbf{A}) + tT(\mathbf{B})$$

for all t.

**Proof.** By Theorem 2.4.8, there is a nonsingular $2 \times 2$ matrix M and a point **P**, so that $T(\mathbf{Q}) = \mathbf{Q}M + \mathbf{P}$ for all **Q**. Now all one has to do is use this formula for T to evaluate both sides of the equation and show that they are equal.

**2.4.12. Theorem.** Affine transformations of the plane preserve the ratio of division.

**Proof.** The theorem is an easy consequence of Propositions 2.4.10 and 2.4.11.

**2.4.13. Theorem.** Affine transformations in the plane multiply area by the absolute value of their determinant.

**Proof.** See [Gans69].

Theorem 2.4.13 points out one of the main intuitions one should have about determinants, namely, that they are intrinsically connected with how transformations expand or shrink area, volume, etc. A precise definition of volume will be given in Chapter 4.

**Definition.** The *equiaffine or equiareal* group is the group of affine transformations with determinant ±1.

Recall our earlier comments how geometric properties are intimately connected to certain groups of transformations. Here are three groups, the "metric" groups, and their associated "metric" properties:

| motions | similarities | equiaffine |
|---|---|---|
| $\Updownarrow$ | $\Updownarrow$ | $\Updownarrow$ |
| distance | angle size | area |

**Definition.** *Affine properties* are properties preserved only by affine transformations (and not by projective transformations, which we will define shortly).

Some affine properties are betweenness, the ratio of division, parallelism, and the concurrence of lines.

**Definition.** Two figures **F** and **F′** are *affinely equivalent* if there is an affine transformation T with $T(\mathbf{F}) = \mathbf{F}'$.

Any two segments, angles, triangles, parallelograms, lines, parabolas, ellipses, and hyperbolas are affinely equivalent. This means that one can use special simple figures to prove things about general figures!

**2.4.14. Example.** To prove that the midpoints of all parallel chords of a parabola **X** are collinear and lie on a line parallel to the axis. See Figure 2.21(a).

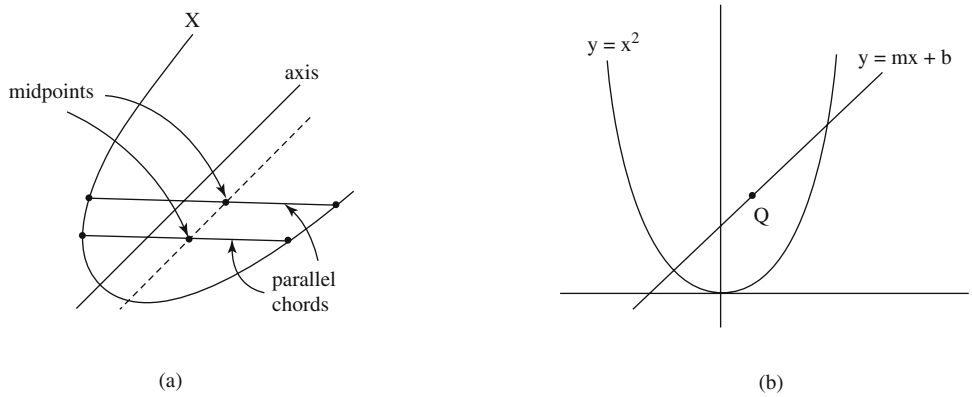



**Figure 2.21.** Midpoints of parallel chords for parabola are parallel to axis.

**Solution.** Since all parabolas are affinely equivalent we may restrict ourselves to the special case of the parabola defined by the equation $y = x^2$ and the family of chords determined by the lines $y = mx + b$, where m is fixed and $b \geq 0$. See Figure 2.21(b). To find the intersection of the lines with the parabola, we must solve the equation $mx + b = x^2$. The two solutions are

$$x_1 = \frac{m + \sqrt{m^2 + 4b}}{2} \quad \text{and} \quad x_2 = \frac{m - \sqrt{m^2 + 4b}}{2}.$$

The midpoint $\mathbf{Q} = (u,v)$ of such a chord is defined by

$$u = \frac{x_1 + x_2}{2} = \frac{m}{2}$$

and

$$v = \frac{mx_1 + b + mx_2 + b}{2} = \frac{m^2 + 2b}{2},$$

which proves the result.

Finally, note that one could have developed affine geometry without first coordinatizing points. We could make points, lines, etc., undefined terms and use axioms to define their properties. This is the *synthetic geometry* approach. Coordinates could be introduced at a later stage. The point is that, in the context of affine geometry, the exact lengths of geometric figures are not important. At most it is **relative** size that counts, that is, the ratios of segments.

## 2.4.1 Parallel Projections

**Definition.** Let $\mathbf{v}$ be a nonzero vector in $\mathbf{R}^n$ and let $\Omega$ be the family of parallel lines with direction vector $\mathbf{v}$. Let $\mathbf{L_p}$ denote the line in $\Omega$ through the point $\mathbf{p}$. If $\mathbf{X}$ is a hyperplane in $\mathbf{R}^n$ not parallel to $\mathbf{v}$, then define a map

$$\pi_\Omega : \mathbf{R}^n \to \mathbf{X}$$

by

$$\pi_\Omega(\mathbf{p}) = \mathbf{L_p} \cap \mathbf{X}.$$

The map $\pi_\Omega$ is called the *parallel projection of* $\boldsymbol{R}^n$ *onto the plane* $\boldsymbol{X}$ *parallel to* $\boldsymbol{v}$. If $\mathbf{v}$ is orthogonal to $\mathbf{X}$, then $\pi_\Omega$ is called the *orthogonal* or *orthographic projection of* $\boldsymbol{R}^n$ *onto the plane* $\boldsymbol{X}$; otherwise, it is called an *oblique parallel projection*. In general, if $\mathbf{X}$ and $\mathbf{Y}$ are any subsets of $\mathbf{R}^n$, then the map that sends $\mathbf{p}$ in $\mathbf{X}$ to $\mathbf{L_p} \cap \mathbf{Y}$ in $\mathbf{Y}$ (wherever it is defined) is called the *parallel projection of* $\boldsymbol{X}$ *to* $\boldsymbol{Y}$.

Figure 2.22 shows a parallel projection of a line $\mathbf{L}$ onto a line $\mathbf{L}'$ and Figure 2.23, a parallel projection of a plane $\mathbf{X}$ onto a plane $\mathbf{X}'$. Note that the ratio of distances is preserved in the case of parallel projections of a line onto another line. What this means is that, referring to Figure 2.22, the ratio

$$\frac{|\mathbf{AB}|}{|\mathbf{A'B'}|}$$

is independent of $\mathbf{A}$ and $\mathbf{B}$. This is **not** the case for parallel projections of one plane onto another. For example, in Figure 2.23 the ratios

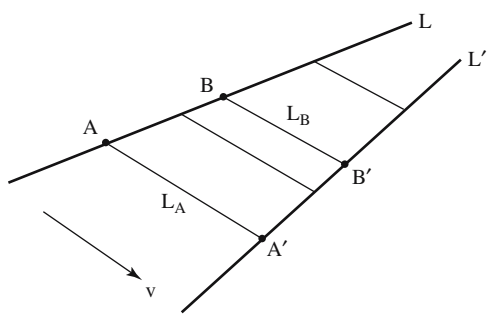



**Figure 2.22.** A parallel projection between lines.

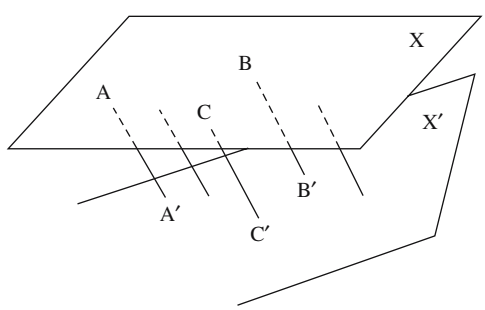



**Figure 2.23.** A parallel projection between planes.

$$\frac{|\mathbf{AB}|}{|\mathbf{A'B'}|} \quad \text{and} \quad \frac{|\mathbf{BC}|}{|\mathbf{B'C'}|}$$

are probably not the same.

**2.4.1.1. Example.** To find the parallel projection T of $\mathbf{R}^3$ onto the plane $\mathbf{X}$ defined by the equation

$$x - 2y + z = 3$$

parallel to $\mathbf{v} = (3,1,1)$.

**Solution.** Clearly, given a point $\mathbf{p}$, if t is chosen so that $\mathbf{p} + t\mathbf{v}$ belongs to $\mathbf{X}$, then $T(\mathbf{p}) = \mathbf{p} + t\mathbf{v}$. Let $\mathbf{p} = (x,y,z)$. Solving

$$(x + 3t) - 2(y + t) + (z + t) = 3$$

for t, gives $t = (1/2)(-x + 2y - z)$. It follows that T is defined by the equations

$$x' = -\frac{1}{2}x + 3y - \frac{3}{2}z + \frac{9}{2}$$

$$y' = -\frac{1}{2}x + 2y - \frac{1}{2}z + \frac{3}{2}$$

$$z' = -\frac{1}{2}x + \quad y + \frac{1}{2}z + \frac{3}{2}.$$

**2.4.1.2. Theorem.** A parallel projection between two hyperplanes in $\mathbf{R}^n$ preserves parallelism, concurrence, betweenness, and the ratio of division.

**Proof.** Easy.

**2.4.1.3. Theorem.** Any map of the plane onto itself that is a composition of parallel projections is an affine map. Conversely, every affine map in the plane is a composite of parallel projections.

**Sketch of proof.** The first statement follows from the fact that lines are preserved. Now let T be an affine map. Assume that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are noncollinear points with $T(\mathbf{A},\mathbf{B},\mathbf{C}) = (\mathbf{A}',\mathbf{B}',\mathbf{C}')$. First, project $\mathbf{R}^2$ to a plane $\mathbf{X}$ that contains $\mathbf{A}$ and $\mathbf{B}$ so that $\mathbf{C}$ gets sent to a point $\mathbf{C}_1$. Next, project $\mathbf{X}$ back to $\mathbf{R}^2$ in such a way as to send $\mathbf{C}_1$ to $\mathbf{C}'$. It follows that the composite of these two projections sends $\mathbf{A}$ to $\mathbf{A}$, $\mathbf{B}$ to $\mathbf{B}$, and $\mathbf{C}$ to $\mathbf{C}'$. Repeat this process on $\mathbf{A}'$ and $\mathbf{B}'$. See Figure 2.24.

The construction in the proof of Theorem 2.4.1.3 shows that any affine map can be realized as a composite of at most six projections.

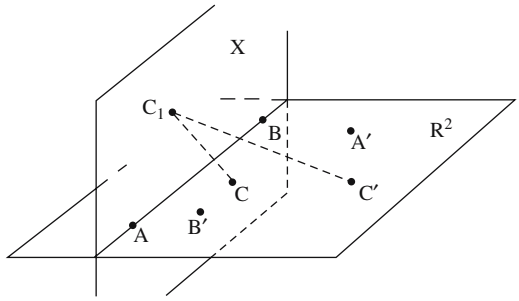



**Figure 2.24.** Affine maps as composites of parallel projections.

## 2.5 Beyond the Plane

Up to now, although some things applied to $\mathbf{R}^n$, most of the details were specifically about transformations in the plane. The fact is that much of what we did generalizes to higher dimensions.

We start with motions of $\mathbf{R}^n$.

**2.5.1. Theorem.** Every motion $M:\mathbf{R}^n \to \mathbf{R}^n$ can be expressed by equations of the form

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + c_1 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + c_2 \\ &\cdot \\ &\cdot \\ x_n' &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + c_n \end{aligned} \tag{2.29}$$

where $A_M = (a_{ij})$ is an orthogonal matrix. Conversely, every such system of equations defines a motion.

**Proof.** The discussion in Section 2.2.8 on frames showed that the theorem is valid for motions in the plane. For the general case, assume without loss of generality that $M(\mathbf{0}) = \mathbf{0}$. The key facts are Theorem 2.2.4.1, which says that M is a linear transformation (and hence is defined by a matrix), and Lemma 2.2.4.3, which says that $M(\mathbf{u}) \bullet M(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v}$, for all vectors $\mathbf{u}$ and $\mathbf{v}$. The rest of the proof simply involves analyzing the conditions $M(\mathbf{e}_i) \bullet M(\mathbf{e}_j) = \mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij}$ and is left as an exercise (Exercise 2.5.1).

In studying motions in the plane we made use of some important special motions, such as translations, rotations, and reflections. Translations already have a general definition. The natural generalization of the definition of a reflection is to replace lines by hyperplanes.

**Definition.** Let $\mathbf{X}$ be a hyperplane in $\mathbf{R}^n$. Define a map $S:\mathbf{R}^n \to \mathbf{R}^n$, called the *reflection about the hyperplane* $\boldsymbol{X}$, as follows: Let $\mathbf{A}$ be a point in $\mathbf{X}$ and let $\mathbf{N}$ be a normal

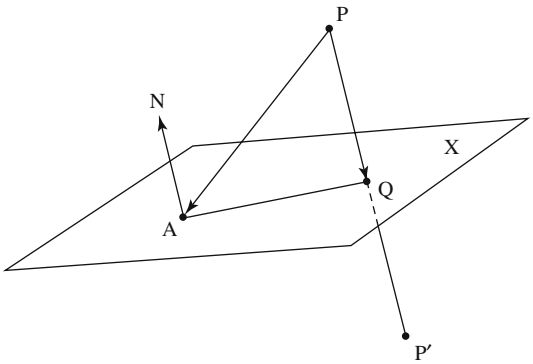



**Figure 2.25.** Defining a reflection in higher dimensions.

vector for **X**. If **P** is any point in $\mathbf{R}^n$, then S(**P**) = **P** + 2**PQ**, where **PQ** is the orthogonal projection of **PA** on **N**. See Figure 2.25.

**2.5.2. Theorem.** Let S be a reflection about a hyperplane **X**.

(1) The definition of S depends only on the hyperplane and not on the point **A** and normal vector **N** that are chosen for it in the definition.
(2) If t is chosen so that **P** + t**N** is the point where the line through **P** with direction vector **N** meets the plane **X**, then S(**P**) = **P** + 2t**N**.
(3) The fixed points of S are just the points of **X**.
(4) If **L′** is a line orthogonal to **X**, then S(**L′**) = **L′**.
(5) Reflections about hyperplanes are motions.

**Proof.** The details of the proof are left as an exercise for the reader because it is essentially the same as the proof of Theorem 2.2.3.1. That proof did not really use the fact that vectors were two-dimensional.

**2.5.3. Example.** To find the reflection S about the plane **X** defined by the equation

$$x - 2y - 2z = 3.$$

**Solution.** Let **A** be any point in **X**. Since **N** = (1,–2,–2) is a normal vector for **X**, if **P** is any point, then it is easy to show that the orthogonal projection of **PA** on **N** is just t**N**, where t is chosen so that **P** + t**N** lies in **X**. Let **P** = (x,y,z). Solving

$$(x + t) - 2(y - 2t) - 2(z - 2t) = 3$$

for t, gives

$$t = (1/9)(-x + 2y + 2z + 3).$$

Since S(**P**) = **P** + 2t**N**, it follows that S has equations

$$x' = \frac{7}{9}x + \frac{4}{9}y + \frac{4}{9}z + \frac{2}{3}$$
$$y' = \frac{4}{9}x + \frac{1}{9}y - \frac{8}{9}z - \frac{4}{3}$$
$$z' = \frac{4}{9}x - \frac{8}{9}y + \frac{1}{9}z - \frac{4}{3}.$$

Generalizing the concept of a rotation is a little less obvious. The simplest way to get a definition is in a roundabout way by defining a rigid motion first and then use the orientation-preserving nature of these maps.

**Definition.** Let M be a motion of $\mathbf{R}^n$ and suppose the equations for M are as shown in Theorem 2.5.1. The motion M is said to be a *rigid motion* if the matrix $(a_{ij})$ is a special orthogonal matrix.

In analogy to the planar case we get

**2.5.4. Theorem.** A rigid motion of $\mathbf{R}^n$ is an orientation-preserving map. Conversely, every orientation-preserving motion of $\mathbf{R}^n$ is a rigid motion.

**Proof.** Exercise.

**Definition.** A rigid motion R of $\mathbf{R}^n$ that fixes some point $\mathbf{p}$ is called a *rotation*. In that case, we say that R is a *rotation about* $\boldsymbol{p}$. The point $\mathbf{p}$ is called a *center* of the rotation.

Is this definition of a rotation really what we want and does it generalize the intuitively simple notion of a rotation in the plane? Theorem 2.2.6.9 certainly shows that the new definition is compatible with the old one.

**2.5.5. Theorem.** (The Principal Axis Theorem) Every rotation R in $\mathbf{R}^3$ is a "rotation about some line." More precisely, with respect to some appropriate coordinate system, R is just the *rotation about the z-axis through an angle* $\theta$, that is, the equations for R in that coordinate system are just

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \\ z' &= \qquad\qquad\quad z. \end{aligned} \tag{2.30}$$

In general, if R is a rotation in $\mathbf{R}^n$, then we can choose a coordinate system with respect to which the $n \times n$ matrix of coefficients in the equation for R has the form

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & & & & & & 0 \\ \sin\theta_1 & \cos\theta_1 & & & & & & \\ & & \ddots & & & & & \\ & & & \cos\theta_k & -\sin\theta_k & & & \\ & & & \sin\theta_k & \cos\theta_k & & & \\ & & & & & \pm 1 & & \\ & & & & & & \ddots & \\ 0 & & & & & & & \pm 1 \end{pmatrix} \tag{2.31}$$

Conversely, every transformation of $\mathbf{R}^n$ whose equation has such a matrix of coefficients is a rotation.

**Proof.** See [Lips68]. Note that rotations about the origin are linear transformations so that one can talk about their associated matrices.

Theorem 2.5.5 suggests that the expression "rotation about a point" is perhaps misleading in higher dimensions. Although it might be better to say "rotation about a line," we shall keep it in order to have a uniform terminology since it makes perfectly good sense in the plane. Actually, we shall see shortly in the next section that one should really talk about directed lines here because the expression "rotation about a line through an angle $\theta$" is **ambiguous**.

The main theorems about motions in $\mathbf{R}^n$ can now be stated. Their proofs are very similar to the proofs of the corresponding theorems about motions in the plane and are omitted.

**2.5.6. Theorem.**

(1) A motion in $\mathbf{R}^n$ is completely determined by what it does to $n + 1$ linearly independent points.
(2) A rigid motion in $\mathbf{R}^n$ is completely determined by what it does to $n$ linearly independent points.
(3) Every motion in $\mathbf{R}^n$ can be described as a composition of a translation, a rotation about the origin, and/or a reflection.
(4) Every rigid motion in $\mathbf{R}^n$ is a composition of a translation and/or a rotation about the origin.

**Proof.** Exercise.

Facts about similarities and affine maps in the plane also generalize to $\mathbf{R}^n$.

**2.5.7. Theorem.** Every similarity transformation can be expressed by equations of the form (2.29) where $(a_{ij}) = (db_{ij})$, $d > 0$, and $(b_{ij})$ is an orthogonal matrix. Conversely, every such system of equations defines a similarity.

**Proof.** Exercise.

**2.5.8. Theorem.**

(1) Every affine transformation T in $\mathbf{R}^n$ can be expressed uniquely in the form

$$T(\mathbf{p}) = \mathbf{p}A + \mathbf{v},$$

where A is an $n \times n$ nonsingular matrix and $\mathbf{v}$ is a fixed vector in $\mathbf{R}^n$. The determinant of A is called the *determinant of the affine transformation*. Conversely, every such equation defines an affine transformation.

(2) An affine transformation is completely specified by its action on $n + 1$ linearly independent points.

(3) The similarity transformations are the angle-preserving affine maps of $\mathbf{R}^n$.

(4) Affine transformations in $\mathbf{R}^n$ multiply volume by the absolute value of their determinant.

**Proof.** Exercise.

**Definition.** A map $T: \mathbf{R}^n \to \mathbf{R}^n$ is said to *preserve barycentric coordinates* if, for all $\mathbf{v}_i \in \mathbf{R}^n$ and real numbers $a_i$,

$$T\left(\sum_{i=0}^{k} a_i \mathbf{v}_i\right) = \sum_{i=0}^{k} a_i T(\mathbf{v}_i) \quad \text{whenever} \quad \sum_{i=0}^{k} a_i = 1. \tag{2.28}$$

**2.5.9. Theorem.** Affine maps in $\mathbf{R}^n$ preserve barycentric coordinates. Conversely, any one-to-one and onto transformation that preserves barycentric coordinates is an affine map.

**Proof.** We prove the first part. Let T be an affine map. By Theorem 2.5.8(1),

$$T(\mathbf{p}) = \mathbf{p}A + \mathbf{v},$$

where A is an $n \times n$ matrix. It follows that

$$\begin{aligned} T\left(\sum_{i=0}^{k} a_i \mathbf{v}_i\right) &= \left(\sum_{i=0}^{k} a_i \mathbf{v}_i\right) A + \mathbf{v} \\ &= \sum_{i=0}^{k} a_i (\mathbf{v}_i A) + \sum_{i=0}^{k} a_i \mathbf{v} \\ &= \sum_{i=0}^{k} a_i (\mathbf{v}_i A + \mathbf{v}) \\ &= \sum_{i=0}^{k} a_i T(\mathbf{v}_i). \end{aligned}$$

**2.5.10. Corollary.** Affine maps in $\mathbf{R}^n$ preserve the ratio of division.

**Proof.** This follows form Proposition 2.4.10 and Theorem 2.5.9.

**Definition.** If $A = (a_{ij})$ is an $n \times n$ nonsingular diagonal matrix, then the transformation $T: \mathbf{R}^n \to \mathbf{R}^n$ defined by

$$T(\mathbf{p}) = \mathbf{p}A$$

is called a (*local*) *scaling transformation*. It is a *global scaling transformation* if all the diagonal elements in A are equal, that is, $a_{11} = a_{22} = \ldots = a_{nn}$.

Note that a scaling transformation is orientation reversing if $|A| < 0$. It will be a similarity if $a_{11} = a_{22} = \ldots = a_{nn} > 0$. It is easy to check that the inverse of a scaling transformation is a scaling transformation.

Facts about parallel projections, such as Theorems 2.4.1.2 and 2.4.1.3, also generalize to $\mathbf{R}^n$. Finally, we generalize frames. See Figure 2.26. These will be especially helpful in higher dimensions as we shall see in Section 2.5.2.

**Definition.** A *frame* in $\mathbf{R}^n$ is a tuple $F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \mathbf{p})$, where $\mathbf{p}$ is a point and the $\mathbf{u}_i$ define an orthonormal basis of $\mathbf{R}^n$. If the ordered basis $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ induces the standard orientation of $\mathbf{R}^n$, then we call the frame an *oriented* frame. The oriented line through the point $\mathbf{p}$ with direction vector $\mathbf{u}_i$ is called the ***$u_i$-axis of the frame F***. In the case of 3-space, the oriented lines through the point $\mathbf{p}$ with direction vectors $\mathbf{u}_1, \mathbf{u}_2$, and $\mathbf{u}_3$ are also called the *x-, y-, and z-axis of F*, respectively. The point $\mathbf{p}$ is called the *origin* of the frame F. $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n, \mathbf{0})$ is called the *standard frame* of $\mathbf{R}^n$. Again, to simplify the notation, we sometimes use $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ to denote the frame $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \mathbf{0})$.

Sometimes one wants to transform frames.

**Definition.** Let $F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \mathbf{p})$ be a frame in $\mathbf{R}^n$. If M is a motion of $\mathbf{R}^n$, define the *transformed frame* M(F) by

$$M(F) = (M(\mathbf{u}_1) - M(\mathbf{0}), M(\mathbf{u}_2) - M(\mathbf{0}), \ldots, M(\mathbf{u}_n) - M(\mathbf{0}), M(\mathbf{p})).$$

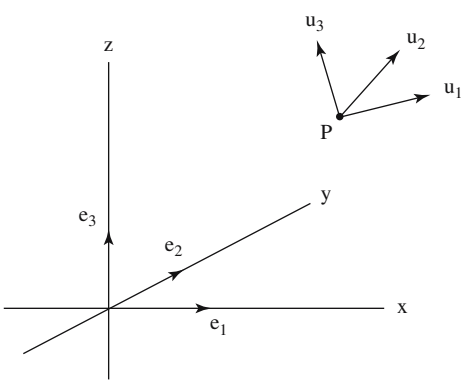



**Figure 2.26.** Frames in $\mathbf{R}^3$.

A transformed frame is clearly a frame. The following generalization to planes is also useful.

**Definition.** A *frame for a k-dimensional plane* $\mathbf{X}$ in $\mathbf{R}^n$ is a tuple $F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{p})$, where $\mathbf{p} \in \mathbf{X}$ and $\mathbf{u}_1, \mathbf{u}_2, \ldots,$ and $\mathbf{u}_k$ are an orthonormal basis for $\mathbf{X}$.

Just like in the two-dimensional case, frames can be thought of as defining a coordinate system for a space and we can coordinatize its points with respect to it. We shall again call the coordinates of a point with respect to a frame the *frame coordinates*. The frame coordinates with respect to the standard frame of $\mathbf{R}^n$ are called *world coordinates*. Frames can also be thought of as motions.

**Definition.** Let $F = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \mathbf{p})$ be a frame in $\mathbf{R}^n$. Define a motion $T_F$, called the *motion defined by F* and usually simply denoted by F, by

$$T_F(\mathbf{q}) = \mathbf{q}\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{pmatrix} + \mathbf{p}. \tag{2.32}$$

Like in the earlier two-dimensional case, using "F" to denote both the frame F and the motion $T_F$ should not cause any confusion since it will always be clear from the context as to whether we are talking about the frame or the map.

Clearly, $T_F$ is a motion. In fact, if T is the translation with translation vector $\mathbf{p}$ and if R is the motion that is the linear transformation with matrix

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n \end{pmatrix},$$

then $T_F = TR$. If the frame F is oriented, then R is a rotation because of our hypothesis that the ordered basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ determines the standard orientation. Furthermore, since we are dealing with motions and orthonormal bases, the matrix of the inverse of R is just the inverse of the matrix for R. It follows that

$$T_F^{-1}(\mathbf{q}) = (\mathbf{q} - \mathbf{p})\left(\mathbf{u}_1^T\ \mathbf{u}_2^T \ldots \mathbf{u}_n^T\right), \tag{2.33}$$

**Note.** In the context of frames as coordinate systems we could have also allowed for *nonorthogonal frames*, that is, tuples $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{p})$ where we only require that the vectors are linearly independent. This would correspond to skew coordinate systems and might be a useful concept in certain situations. The only complication is that finding inverses is no longer as trivial as it is for an orthogonal matrix. Other than that, though, there is nothing new.

### 2.5.1 Motions in 3-Space

In this section we look at the mechanics of transforming objects in 3-space. This may not seem as easy as it was in the plane, but if we break the general problem into a sequence of simple primitive ones, then it will become easy again.

Rigid motions are composites of rotations and/or translations. Is is useful to have some alternate characterizations of rotations. The first characterization comes from the Principal Axis Theorem (Theorem 2.5.5), which says that every rotation is a rotation about an axis. Before we can make use of this way of looking at a rotation we must resolve an ambiguity that we alluded to in a comment immediately following Theorem 2.5.5. Suppose that **v** is a direction vector for the axis. If we consider a plane orthogonal to the axis of the rotation, the notion of counterclockwise for this plane, which is what is normally used to define the positive direction for an angle, will depend on whether we are looking down on this plane from a point on the axis in the **v** or –**v** direction. The only way that this ambiguity in the expression "a rotation about a line through a given angle" can be avoided is by requiring the line to be oriented.

**The axis-angle representation of a rotation:** Here we represent a rotation by a triple $(\mathbf{p},\mathbf{u},\theta)$, where the point **p** and **unit** (direction) vector **u** specify the axis and $\theta$ is the angle of rotation determined according to the following rule:

**The rotation orientation rule:** Think of **u** as being the z-axis for a coordinate system at **p**. Stand at **p** + **u** and look towards the "origin" **p**. The counterclockwise direction in the "x-y plane" of this coordinate system will then determine the positive direction for an angle. See Figure 2.27. More precisely, choose vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ so that $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u})$ forms an orthonormal basis for $\mathbf{R}^3$ that induces the standard orientation. Then $(\mathbf{u}_1,\mathbf{u}_2)$ induces the desired orientation on the x-y plane of the coordinate system from which "clockwise" and "counterclockwise" are determined. The rule can also be expressed in terms of the so-called "right-hand rule," that is, if one lets the thumb of one's right hand point in the direction of **u**, then the curl of the fingers will specify the positive direction of angles. See Figure 2.27 again.

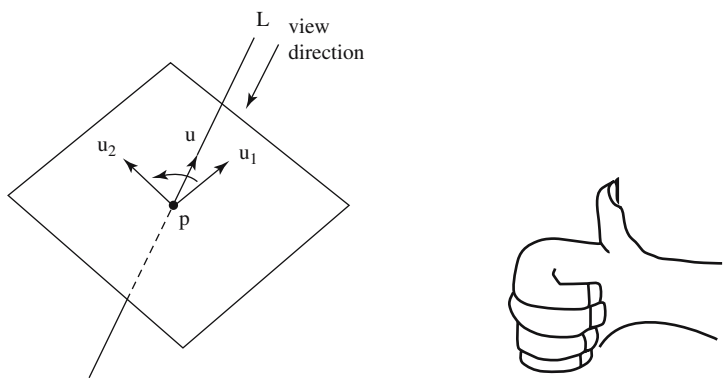



**Figure 2.27.** The orientation of a rotation angle using the right-hand rule.

Note that the direction vector **u** that is chosen for the axis matters because $(\mathbf{p},-\mathbf{u},\theta) = (\mathbf{p},\mathbf{u},-\theta)$.

**Definition.** The rotation defined by the triple $(\mathbf{p},\mathbf{u},\theta)$ is called the *rotation about the directed line* defined by the point **p** and direction vector **u** *through the angle* $\theta$.

We can represent a rotation more compactly by incorporating the angle of the rotation in the length of the direction vector for the axis.

**A compact axis-angle representation of a rotation:** Here a pair $(\mathbf{p},\mathbf{v})$ represents the rotation whose axis-angle representation is $(\mathbf{p},\mathbf{v}/|\mathbf{v}|,|\mathbf{v}|)$.

The next two characterizations of rotations are in terms of rotating about coordinate axes. Fortunately, when it comes to rotations we only need to know the equations of the rotations about the coordinate axes by heart. It is therefore worthwhile to summarize those before moving on since the equations for all other rotations can be derived from them.

**The equations and matrices for the rotations about the coordinate axes:**

| rotation about x-axis | rotation about y-axis | rotation about z-axis |
|---|---|---|
| $x' = x$ | $x' = x\cos\theta + z\sin\theta$ | $x' = x\cos\theta - y\sin\theta$ |
| $y' = y\cos\theta - z\sin\theta$ | $y' = y$ | $y' = x\sin\theta + z\cos\theta$ |
| $z' = y\sin\theta + z\cos\theta$ | $z' = -x\sin\theta + z\cos\theta$ | $z' = z$ (2.34) |

$$\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{2.35}$$

Note that the minus sign in the equation and matrix of a rotation about the y-axis is different from the other rotations. The reason is that we are expressing things in world coordinates and when looking down the y-axis, the x-axis is pointing to the left which is the wrong direction because angles are oriented according to the basis $(-\mathbf{e}_1,\mathbf{e}_3)$.

**2.5.1.1. Theorem.** Consider a coordinate system specified by a frame $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{p})$. If R is a rotation about **p**, then R is the composite of a rotation $R_1$ about the x-axis of F through an angle $\alpha$, a rotation $R_2$ about the y-axis of F through an angle $\beta$, and finally a rotation $R_3$ about the z-axis of F through an angle $\tau$.

**Proof.** Assume first that we are rotating about the origin and F is the standard frame. With this hypothesis, R is a linear transformation and has matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{2.36}$$

The rotations $R_1$, $R_2$, and $R_3$ would also be linear transformations and we know their matrices from (2.35). Therefore, using the abbreviations

$$c\tau = \cos\tau,\ s\tau = \sin\tau,\ c\beta = \cos\beta,\ s\beta = \sin\beta,\ c\alpha = \cos\alpha, \quad \text{and} \quad s\alpha = \sin\alpha,$$

the matrix for $R_3R_2R_1$ would be

$$\begin{pmatrix} c\tau\, c\beta & s\tau\, c\beta & -s\beta \\ c\tau\, s\beta\, s\alpha - s\tau\, c\alpha & c\tau\, s\beta\, s\alpha + c\tau\, c\alpha & c\beta\, s\alpha \\ c\tau\, s\beta\, c\alpha + s\tau\, s\alpha & c\tau\, s\beta\, c\alpha - c\tau\, s\alpha & c\beta\, c\alpha \end{pmatrix}. \tag{2.37}$$

To prove the theorem, all we have to do is set the matrices in (2.36) and (2.37) equal to each other and solve for α, β, and τ. This is not hard using the first row and last column. If $c\beta \neq 0$, then

$$\alpha = \operatorname{atan2}(a_{23}, a_{33}), \tag{2.38a}$$

$$\beta = \operatorname{atan2}\left(a_{13}, \sqrt{a_{11}^2 + a_{12}^2}\right), \tag{2.38b}$$

$$\tau = \operatorname{atan2}(a_{12}, a_{11}), \tag{2.38c}$$

where atan2(y,x) is basically the arctangent $\tan^{-1}(y/x)$, except that the sign of both x and y are used to determine into which quadrant the angle falls. See Appendix A for a precise definition.

The case of an arbitrary frame F is an easy consequence of this case that we leave as an exercise. The theorem is proved.

**Note.** In the proof of Theorem 2.5.1.1, only one choice had to be made in defining α, β, and τ and that was the choice of the positive square root of the sum of squares of $a_{11}$ and $a_{21}$. This amounts to restricting β to lying in the interval $[-\pi/2, \pi/2]$. With this restriction, the α, β, and τ are uniquely determined for R.

**Definition.** Given a rotation R, the angles α, β, and τ in Theorem 2.5.1.1 are called the *roll, pitch, and yaw angles of R*, respectively. The tuple $[\alpha,\beta,\tau,\mathbf{p}]$ is called a *roll-pitch-yaw representation of the rotation R* (with respect to the frame F). $[\alpha,\beta,\tau]$ will denote the roll-pitch-yaw representation in the case where **p** is the origin.

The terminology of "roll," "pitch," and "yaw" comes from aviation and navigation. Roll is the twisting motion about the lengthwise axis of a ship or airplane. Pitch is the dipping or rising motion of the bow or nose. Yaw is the side-to-side twisting motion in its horizontal plane about a vertical axis. The note following the theorem above shows that the roll, pitch, and yaw angles are unique if the pitch angle lies in the interval $[-\pi/2, \pi/2]$.

Instead of rotating about the axes of a fixed coordinate system as is done in the case of the roll-pitch-yaw representation of a rotation we can do our rotations about axes in each successive new coordinate system. The choice of axes is up to us.

**2.5.1.2. Theorem.** Consider a coordinate system specified by a frame F = $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{p})$. If R is a rotation about **p**, then R is the composite of a rotation $S_1$ about

the x-axis of F through an angle $\alpha$, a rotation $S_2$ about the y-axis of $S_1(F)$ through an angle $\beta$, and finally a rotation about the z-axis of $S_2(S_1(F))$ through an angle $\tau$.

**Proof.** We shall again only consider the case where F is the standard frame and **p** is the origin. If $R_1$, $R_2$, and $R_3$ are the rotations about the standard coordinate axes defined in Theorem 2.5.1.1, then

$$\begin{aligned} S_1 &= R_1, \\ S_2 &= S_1 R_2 S_1^{-1}, \\ S_3 &= S_2 S_1 R_3 (S_2 S_1)^{-1}, \end{aligned}$$

so that $S_3S_2S_1 = R_1R_2R_3$. Note that this composition is in the opposite order of the composition of the maps in Theorem 2.5.1.1, but the matrix for $S_3S_2S_1$ will be similar to the one shown in (2.37). Therefore we can set this matrix equal to the matrix for R and solve for the angles just like in Theorem 2.5.1.1.

**Definition.** Given a rotation R, the angles $\alpha$, $\beta$, and $\tau$ in Theorem 2.5.1.2 are called the *X-Y-Z Euler angles of R*. The tuple $[\alpha,\beta,\tau,\mathbf{p}]$ is called an *Euler angle representation of the rotation R* (with respect to the frame F). $[\alpha,\beta,\tau]$ will denote the Euler angle representation in the case where **p** is the origin.

The term Euler angles is also used in the case of any other choice of axes. For example, if one were to rotate about the z-, y-, and z-axis, then one would get the *Z-Y-Z Euler angles* for a rotation, and so on. We shall only look at Euler angles in the X-Y-Z case. The others are similar. The proofs of Theorems 2.5.1.1 and 2.5.1.2 show us how to compute the roll, pitch, yaw or Euler angles of a rotation. We shall not pursue the subject here. These angles are often used to describe motions in robotics.

Let us return to the main subject matter of this section, which is how to derive equations for rigid motions. Since translations are trivial, we now work through an example to show how one typically computes equations for geometrically defined rotations. The idea is to express an arbitrary rotation in terms of rotations about the x-, y-, and z-axis.

**2.5.1.3. Example.** To show that any nonzero vector **v** can be rotated into one of the coordinate axes by a composition of two rotations about coordinate axes.

**Solution.** We sketch the construction in case we want to rotate **v** into the z-axis. Rotating into the x- or y-axis would be done in a similar way. Let **w** by the orthogonal projection of **v** onto the y-z plane. See Figure 2.28(a). A rotation $R_1$ about the x-axis through an angle $\alpha$, where $\alpha$ is the angle that **w** makes with the z-axis (or $\mathbf{e}_3$), will move **v** into a vector $\mathbf{v}'$ in the x-z plane. See Figure 2.28(b). A second rotation $R_2$ about the y-axis through the angle $\beta$, where $\beta$ is the angle that $\mathbf{v}'$ makes with the z-axis, will rotate $\mathbf{v}'$ into the z-axis. The composition $R_2R_1$ then does what we want, namely, move **v** into the z-axis.

**2.5.1.4. Example.** To find the rotation R that rotates the plane **X** defined by

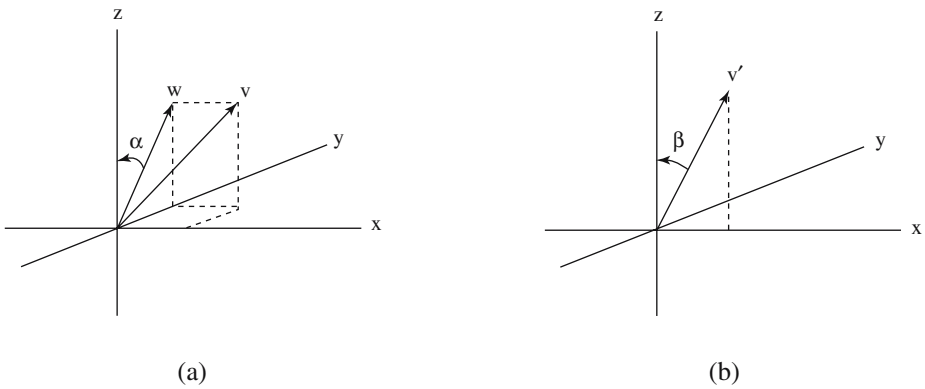



**Figure 2.28.** Rotating a vector into the z-axis.

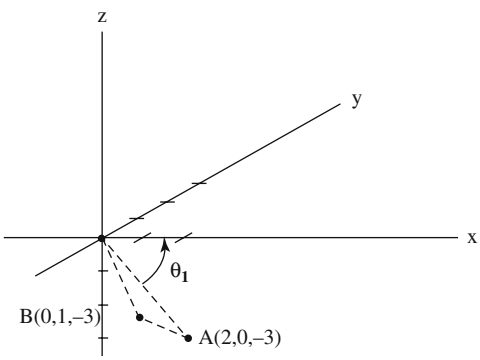



**Figure 2.29.** Example 2.5.1.4.

$$\frac{3}{2}x + 3y + z = 0$$

to the x-y plane.

**Solution.** See Figure 2.29. The idea is to express R as a composite of our basic rotations about the coordinate axes. Recall that motions map planes to planes and so to define R we only need to define an R that does the right thing on three noncollinear points in **X**. We shall use **O** = (0,0,0), **A** = (2,0,–3), and **B** = (0,1,–3) and deal with the points **A** and **B** one at a time. Let $R_1$ be the rotation about the y-axis through an angle $\theta_1$ defined by

$$\cos\theta_1 = \frac{2}{\sqrt{13}} \quad \text{and} \quad \sin\theta_1 = \frac{3}{\sqrt{13}}.$$

The matrix for $R_1$ is

$$\begin{pmatrix} \frac{2}{\sqrt{13}} & 0 & \frac{3}{\sqrt{13}} \\ 0 & 1 & 0 \\ -\frac{3}{\sqrt{13}} & 0 & \frac{2}{\sqrt{13}} \end{pmatrix}.$$

It follows that

$$\mathbf{B}_1 = \mathbf{R}_1(\mathbf{B}) = \left(\frac{9}{\sqrt{13}}, 1, -\frac{6}{\sqrt{13}}\right).$$

Let $R_2$ be the rotation about the x-axis through an angle $\theta_2$ where

$$\cos\theta_2 = \frac{\sqrt{13}}{7} \quad \text{and} \quad \sin\theta_2 = \frac{6}{7}.$$

$R_2$ will move $\mathbf{B}_1$ to the x-y plane and leave $\mathbf{A}_1$ fixed. Finally, $R = R_2R_1$ will be the rotation we are looking for because R leaves the origin fixed and maps the points **A** and **B** to the x-y plane. The matrix for $R_2$ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{13}}{7} & \frac{6}{7} \\ 0 & -\frac{6}{7} & \frac{\sqrt{13}}{7} \end{pmatrix},$$

and so the matrix for R is

$$\begin{pmatrix} \frac{2}{\sqrt{13}} & -\frac{18}{7\sqrt{13}} & \frac{3}{7} \\ 0 & \frac{\sqrt{13}}{7} & \frac{6}{7} \\ -\frac{3}{\sqrt{13}} & -\frac{12}{7\sqrt{13}} & \frac{2}{7} \end{pmatrix}.$$

Before leaving this problem, let us look at another possible question. What is the equation for the plane $\mathbf{X}_1 = R_1(\mathbf{X})$? Note that, by definition, $(x_1,y_1,z_1)$ belongs to $\mathbf{X}_1$ if and only if $R_1^{-1}(x_1,y_1,z_1)$ belongs to **X**. Therefore, since the matrix for $R_1^{-1}$ is

$$\begin{pmatrix} \frac{2}{\sqrt{13}} & 0 & -\frac{3}{\sqrt{13}} \\ 0 & 1 & 0 \\ \frac{3}{\sqrt{13}} & 0 & \frac{2}{\sqrt{13}} \end{pmatrix}$$

substituting into the equation for **X** gives that

$$\frac{3}{2}\left[\frac{2}{\sqrt{13}}x+\frac{3}{\sqrt{13}}z\right]+3y+\left[-\frac{3}{\sqrt{13}}x+\frac{2}{\sqrt{13}}z\right]=0,$$

that is,

$$3\sqrt{13}y+\frac{13}{2}z=0,$$

is the equation for $\mathbf{X}_1$. This shows that $\mathbf{X}_1$ contains the x-axis and is orthogonal to the y-z plane, justifying our construction. This finishes Example 2.5.1.4.

## 2.5.2 Frames Revisited

The last section described what might be called the geometric approach to defining motions in $\mathbf{R}^3$. Some of the computations got rather complicated. The power of frames comes from their ability to define a motion M in terms of an orthonormal basis, which is typically easier to define than the rotations and reflections that might describe M if we were to use the approach from the last section. We saw some of this in Section 2.2.8, but it is especially going to pay off here. As our first example we redo Example 2.5.1.4.

**2.5.2.1. Example.** To find the rotation R that rotates the plane **X** defined by

$$\frac{3}{2}x+3y+z=0$$

to the x-y plane.

**Solution.** We use the same notation as in Example 2.5.1.4. See Figure 2.29. Applying the Gram-Schmidt algorithm to the basis **A**(2,0,–3) and **B**(0,1,–3) for **X** gives us an orthonormal basis

$$\mathbf{u}_1=\frac{1}{\sqrt{13}}(2,0,-3),\quad \text{and}\quad \mathbf{u}_2=\frac{1}{7\sqrt{13}}(-18,13,-12).$$

The equation for **X** tells us that **n** = (3/2,3,1) is a normal vector for the plane. Let

$$\mathbf{u}_3=\frac{\mathbf{n}}{|\mathbf{n}|}=\frac{1}{7}(3,6,2),$$

and consider the frame F = ($\mathbf{u}_1$,$\mathbf{u}_2$,$\mathbf{u}_3$). The rotation R defined by $F^{-1}$ then solves the problem. The matrix for R is the same one as we got before, namely,

$$\begin{pmatrix} \frac{2}{\sqrt{13}} & -\frac{18}{7\sqrt{13}} & \frac{3}{7} \\ 0 & \frac{\sqrt{13}}{7} & \frac{6}{7} \\ -\frac{3}{\sqrt{13}} & -\frac{12}{7\sqrt{13}} & \frac{2}{7} \end{pmatrix}.$$

Actually, the fact that we got the same answer is accidental since the problem is under-constrained and there are many rotations that rotate **X** to the x-y plane.

**2.5.2.2. Example.** To find the rotation R which rotates the plane **X** defined by the equation y − z = 0 to the x-y plane.

**Solution.** By inspection it is clear that the vectors

$$\mathbf{u}_1 = (1,0,0), \quad \mathbf{u}_2 = \left(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \text{and} \quad \mathbf{u}_3 = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

are an orthonormal basis of $\mathbf{R}^3$ with $\mathbf{u}_1$ and $\mathbf{u}_2$ a basis for **X**. See Figure 2.30. Define the orthogonal matrix A by

$$A = \left(\mathbf{u}_1^T\ \mathbf{u}_2^T\ \mathbf{u}_3^T\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

A is matrix for the rotation R we are looking for. It is easy to check that $\mathbf{u}_i A = \mathbf{e}_i$. Note that

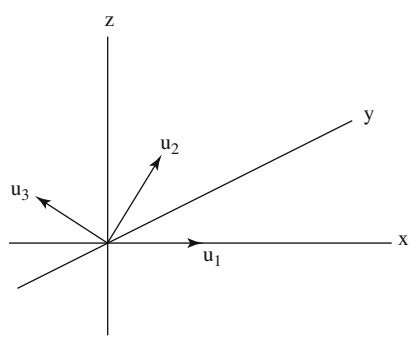



**Figure 2.30.** Example 2.5.2.2.

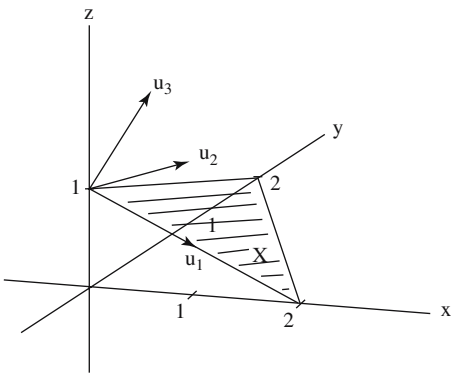



**Figure 2.31.** Example 2.5.2.3.

$$A^{-1} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

and that $\mathbf{e}_i A^{-1} = \mathbf{u}_i$.

The approach used in the last two examples generalizes.

**2.5.2.3. Example.** To find a rigid motion M that moves the origin to the point $\mathbf{p} = (0,0,1)$ and the x-y plane to the plane $\mathbf{X}$ defined by $x + y + 2z = 2$. See Figure 2.31.

**Solution.** All we have to do is to find an orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$, so that $\mathbf{u}_1$ and $\mathbf{u}_2$ are a basis for $\mathbf{X}$. Then the motion defined by the frame $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{p})$ will do the job. The equation for $\mathbf{X}$ tells us that $\mathbf{n} = (1,1,2)$ is a normal vector for $\mathbf{X}$. There are many ways to find a basis for $\mathbf{X}$. Clearly, $(2,0,-1)$ is a vector orthogonal to $\mathbf{n}$. Therefore, let

$$\mathbf{u}_3 = \frac{\mathbf{n}}{|\mathbf{n}|} = \frac{1}{\sqrt{6}}(1,1,2), \quad \mathbf{u}_1 = \frac{1}{\sqrt{5}}(2,0,-1), \quad \text{and} \quad \mathbf{u}_2 = \mathbf{u}_3 \times \mathbf{u}_1 = \frac{1}{\sqrt{30}}(-1,5,-2).$$

**2.5.2.4. Example.** To find a rigid motion M which moves the point $\mathbf{p} = (0,0,1)$ to the point $\mathbf{q} = (0,-1,0)$ and the plane $\mathbf{X}$ defined by $x + y + 2z = 2$ to the plane $\mathbf{Y}$ defined by $-x - y + z = 1$.

**Solution.** Let F be the frame defined in Example 2.5.2.3. The motion $F^{-1}$ will map the plane $\mathbf{X}$ to the x-y plane and $\mathbf{p}$ to the origin. We simply need to define a frame $G = (\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{q})$ that will send the x-y plane to the plane $\mathbf{Y}$ and the origin to $\mathbf{q}$ and set $M = GF^{-1}$. This is another problem like the one in Example 2.5.2.3. Let

$$\mathbf{w}_3 = \frac{1}{\sqrt{3}}(-1,-1,1), \quad \mathbf{w}_1 = \frac{1}{\sqrt{2}}(1,0,1), \quad \text{and} \quad \mathbf{w}_2 = \mathbf{w}_3 \times \mathbf{w}_1 = \frac{1}{\sqrt{6}}(-1,2,1).$$

It follows that

$$(x',y',z') = M(x,y,z) = (x,y,z-1)\begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} & \frac{2}{\sqrt{6}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} + (0,-1,0)$$

For emphasis, we note that the solutions to the last four examples are not unique. For example, to define M in Example 2.5.2.4 we could have picked **any** orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2)$ for $\mathbf{X}$. It is essential however that one picks an **orthonormal** basis, namely, a basis that consists of **unit** vectors that are **mutually orthogonal**. If either of these conditions does not hold, then answers will be wrong.

## 2.6 Exercises

### Section 2.2

2.2.1. Prove Theorem 2.2.1(2) and (3).

2.2.2. Prove that motions send triangles to triangles.

2.2.3. Prove that motions send rays to rays.

### Section 2.2.1

2.2.1.1. Prove Theorem 2.2.1.1.

2.2.1.2. Prove Proposition 2.2.1.2(2).

### Section 2.2.2

2.2.2.1. Prove Proposition 2.2.2.7(1).

2.2.2.2. Find the rotation about the point (1,2) through an angle of $\pi/6$.

2.2.2.3. Let R be a rotation about the origin through an angle $\pi/3$. Let $\mathbf{L}$ be the line determined by the two points (2,4) and $(4,4-2/\sqrt{3})$. Show by direct computation that the angles that $\mathbf{L}$ and $\mathbf{L}' = R(\mathbf{L})$ make with the x-axis differ by $\pi/3$.

2.2.2.4. Find the rotation R about the point (2,3) that sends (6,3) to $(4,3+\sqrt{3})$.

2.2.2.5. Let R be the rotation about (–1,2) through an angle of $-\pi/6$. Let $\mathbf{L}$ be the line determined by the points (2,4) and (5,1). Find the equation for $\mathbf{L}' = R(\mathbf{L})$.

2.2.2.6. If R is the rotation about the origin through an angle of $\pi/3$ degrees and if T is the translation with translation vector (–1,2), then find the equation for RT and describe the map in geometric terms as precisely as possible.

**Section 2.2.3**

2.2.3.1. Prove Theorem 2.2.3.1.

2.2.3.2. Find the reflection S about the line **L** defined by the equation $2x + y = 2$. Find S(–5,0) and S(0,4).

(a) Solve the problem using the definition like in Examples 2.2.3.2 and 2.2.3.3 but check your answer using Proposition 2.2.3.4.
(b) Solve the problem using the reductive method like in Example 2.2.3.7.

2.2.3.3. Let S be the reflection about the line **L** through the points (2,3) and (4,1). Find the equation for S like you were asked to do in Exercise 2.2.3.2.

2.2.3.4. Suppose that $R_1$ and $R_2$ are reflections about the lines $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively. Let $R = R_2R_1$.

(a) If $\mathbf{L}_1$ and $\mathbf{L}_2$ intersect in a point **A**, show that R is a rotation about **A**. Find the relationship between the angle of this rotation and the angle between the two lines.
(b) If $\mathbf{L}_1$ is parallel to $\mathbf{L}_2$, show that R is a translation.

**Section 2.2.4**

2.2.4.1. Prove Theorem 2.2.4.6.

**Section 2.2.7**

2.2.7.1. Prove the converse to Theorem 2.2.7.1.

2.2.7.2. Which of the transformations M below are motions? Explain your answers. In particular, express those that are in the form of a composite of a translation, rotation, and/or reflection:

(a) M: $x' = \frac{3}{5}x + \frac{4}{5}y - 6$, $y = -\frac{4}{5}x + \frac{3}{5}y + 1$

(b) M: $x' = \frac{\sqrt{5}}{2}x + \frac{1}{2}y + 1$, $y' = \frac{1}{2}x - \frac{\sqrt{5}}{2}y + 3$

(c) M: $x' = \frac{3}{2}x + \frac{1}{2}y$, $y' = \frac{1}{2}x - \frac{1}{2}y - 2$

(d) M: $x' = \frac{\sqrt{8}}{3}x + \frac{1}{3}y + 7$, $y' = \frac{1}{3}x - \frac{\sqrt{8}}{3}y$

2.2.7.3. (a) Find the equation for the rigid motion M which sends **A**(2,–1), **B**(4,1) to **A′**(–3,3), **B′**(–1,1), respectively. Use only translations, rotations, and/or reflections.
(b) Find the equation of **another** motion that sends **A** and **B** to **A′** and **B′**, respectively.

2.2.7.4. Find the equation of the motion M that sends **A**(–2,1), **B**(0,2), and **C**(–2,4) to **A′**(4,0), **B′**(6,–1), and **C′**(4,–3), respectively. Use only translations, rotations, and/or reflections.

2.2.7.5. Show that any motion of the form

$$\begin{aligned} M:\ x' &= ax + by \\ y' &= bx - ay, \quad \text{where} \quad a^2 + b^2 = 1, \end{aligned}$$

is a reflection about a line through the origin.

2.2.7.6. Explain why

$$\begin{aligned} M:\ x' &= -\frac{4}{5}x + \frac{3}{5}y + 6 \\ y' &= \frac{3}{5}x + \frac{4}{5}y + 1 \end{aligned}$$

is an orientation reversing motion that is **not** a reflection. On the other hand, M has a fixed line. Find it.

2.2.7.7. Show that every orientation reversing motion is a composite of a rigid motion and a single reflection.

2.2.7.8. Prove the following:

(a) Every motion can be expressed as a composition of at most three reflections.
(b) Every motion with one fixed point is the composite of at most two reflections.

**Section 2.2.8**

2.2.8.1. Use frames to find a motion that sends the line **L** through **A**(2,1) and **B**(3,3) to the x-axis and the point **A** to the origin.

2.2.8.2. Use frames to find a motion which sends the line $2x + 3y = 5$ to the line $x - 2y = 3$.

2.2.8.3. Solve Exercise 2.2.7.3 using frames.

2.2.8.4. Use frames to find the equations of the motion that sends the points **A**(–1,3), **B**(0,1), and **C**(–2,1) to **A′**(3,2), **B′**(2,0), and **C′**(1,2), respectively.

2.2.8.5. Consider the lines

$$\mathbf{L}_1:\ x + 3y = 9 \quad \text{and} \quad \mathbf{L}_2:\ 3x - y = 7.$$

Assuming that the lines are oriented to the right, find the transformation that converts from world coordinates to the coordinate system where $\mathbf{L}_1$ and $\mathbf{L}_2$ are the x- and y-axis, respectively.

**Section 2.3**

2.3.1. Find the equations of the similarity S that sends the points **A**(–1,3), **B**(0,1), and **C**(–2,1) to **A′**(0,6), **B′**(2,2), and **C′**(4,6), respectively.

**Section 2.4**

2.4.1. Prove that the inverse of the shear

$$\begin{aligned} x' &= x \\ y' &= cx + y \end{aligned}$$

is a shear and find its equations.

2.4.2. Find the affine map that sends **A**(1,0), **B**(0,3), and **C**(4,2) to **A′**(–1,2), **B′**(0,1), and **C′**(1,3), respectively.

**Section 2.4.1**

2.4.1.1. Let **X** be the plane defined by x – 2y + z = 1.

(a) Define the orthogonal projection of $\mathbf{R}^3$ onto **X**.
(b) Define the parallel projection of $\mathbf{R}^3$ onto **X** parallel to **v** = (1,0,2).

**Section 2.5**

2.5.1. Fill in the missing details in the proof of Theorem 2.5.1.

2.5.2. Using the definition, find the equation of the reflection S about the plane x – 2y + 2z = 1.

**Section 2.5.1**

2.5.1.1. Show that the following motion is a rotation and find its axis and angle of rotation:

$$\begin{aligned} x' &= \frac{1}{6}x + \left(\frac{2}{\sqrt{6}} + \frac{1}{6}\right)y + \left(-\frac{1}{\sqrt{6}} + \frac{1}{3}\right)z \\ y' &= \left(-\frac{2}{\sqrt{6}} + \frac{1}{6}\right)x + \frac{1}{6}y + \left(\frac{1}{\sqrt{6}} + \frac{1}{3}\right)z \\ z' &= \left(\frac{1}{\sqrt{6}} + \frac{1}{3}\right)x + \left(-\frac{1}{\sqrt{6}} + \frac{1}{3}\right)y + \frac{2}{3}z \end{aligned}$$

2.5.1.2. Using translations and rotations about the coordinate axes, find the equation of a rigid motion that sends the plane **X** defined by 2x – 3y + 2z = 1 to the x-y plane.

2.5.1.3. Given a unit cube with one corner at (0,0,0) and the opposite corner at (1,1,1), derive the transformations necessary to rotate the cube by θ degrees about the main diagonal (from (0,0,0) to (1,1,1)) in the counterclockwise direction when looking along the diagonal toward the origin. Use rotations about the coordinate axes.

2.5.1.4. Show that the motion

$$\begin{aligned} x' &= \frac{1}{2}x + \frac{1}{\sqrt{2}}y + \frac{1}{2}z - 2 \\ y' &= -\frac{1}{2}x + \qquad \frac{1}{\sqrt{2}}z \\ z' &= \frac{1}{2}x - \frac{1}{\sqrt{2}}y + \frac{1}{2}z - 2 \end{aligned}$$

is a screw motion, where

**Definition.** A motion in $\mathbf{R}^3$ of the form RT, where R is a rotation that is not the identity map and T is a translation with translation vector parallel to the line that R is rotating about, is called a *screw motion*.

2.5.1.5. Find the equations for the rotation whose roll-pitch-yaw representation is $(\pi/2,\pi/3,\pi)$.

2.5.1.6. Find the equations for the rotation whose X-Y-Z Euler angle representation is $(\pi/2,\pi/3,\pi)$.

**Section 2.5.2**

2.5.2.1. Solve Exercise 2.5.1.2 using frames.

2.5.2.2. Solve Exercise 2.5.1.3 using frames.

2.5.2.3. Use frames to find the equations of the motion that sends the points $\mathbf{A}(1,0,0)$, $\mathbf{B}(0,1,0)$, $\mathbf{C}(0,0,1)$, and $\mathbf{D}(1,2,1)$ to

$$\mathbf{A}'\left(0,1,-\frac{2}{\sqrt{13}}\right),\quad \mathbf{B}'\left(-\frac{2}{\sqrt{13}},2,\frac{1}{\sqrt{13}}\right),\quad \mathbf{C}'\left(\frac{1}{\sqrt{13}},1,\frac{3}{\sqrt{13}}\right),\quad \text{and}\quad \mathbf{D}'\left(\frac{3}{\sqrt{13}},3,0\right),$$

respectively.

CHAPTER 3

# Projective Geometry

## 3.1 Overview

The last chapter outlined some of the basic elements of affine geometry. This chapter looks at projective geometry. Some general references that look at the subject in more detail than we are able to here are [Ayre67], [Gans69], and [PenP86].

Like in the last chapter, we shall start with dimension two (Sections 3.2–3.4) and only get to higher dimensions in Section 3.5. In order to motivate the transition from affine geometry to projective geometry we begin by studying projective transformations in affine space. Section 3.2 starts off by looking at central projections and leads up to a definition of a projective transformation of the plane. We shall quickly see that, in contrast to affine geometry, we have to deal with certain exceptional cases that make the statement of definitions and theorems rather awkward. Mathematicians do not like having to deal with results on a case-by-case basis. Furthermore, the existence of special cases often is a sign that one does not have a complete understanding of what is going on and that there is still some underlying general principle left to be discovered. In fact, it will become clear that Euclidean affine space is not the appropriate space to look at when one wants to study projective transformations and that one should really look at a larger space called projective space. This will allow us to deal with our new geometric problems in a uniform way.

Projective space itself can be introduced in different ways. One can start with a synthetic and axiomatic point of view or one using coordinates. Lack of space prevents us from discussing both approaches and so we choose the latter because it is more practical. In Section 3.3 we introduce homogeneous coordinates after a new look at points and lines that motivates the point of view that projective space is a natural coordinate system extension of Euclidean space. This leads to a definition and discussion of the projective plane $\mathbf{P}^2$ in Section 3.4. Some of its important analytic properties are described in Section 3.4.1. Sections 3.4.2 and 3.4.3 define projective transformations of $\mathbf{P}^2$ and show how affine transformations are just special cases if one uses homogeneous coordinates. We then generalize to higher dimensions in Section 3.5. The important special case of 3-dimensional projective transformations is considered in Section 3.5.1. Next we study conics (Sections 3.6 and 3.6.1) and quadric surfaces (Section 3.7).

We finish the chapter with several special topics. Section 3.8 discusses a generalization of the usual central projection. Section 3.9 describes the beautiful theorem of Pascal and some applications. The last topic of the chapter is the stereographic projection. Section 3.10 describes some of its main properties.

## 3.2 Central Projections and Perspectivities

**Definition.** Let $\mathbf{O}$ be a fixed point of $\mathbf{R}^n$. For every point $\mathbf{p}$ of $\mathbf{R}^n$ distinct from $\mathbf{O}$, let $\mathbf{L_p}$ denote the line through $\mathbf{O}$ and $\mathbf{p}$. If $\mathbf{Y}$ is a hyperplane in $\mathbf{R}^n$ not containing $\mathbf{O}$, then define a map

$$\pi_o : \mathbf{R}^n \to \mathbf{Y}$$

by

$$\begin{aligned}\pi_{\mathbf{O}}(\mathbf{p}) &= \mathbf{L_p} \cap \mathbf{Y}, \text{ if } \mathbf{L_p} \text{ intersects } \mathbf{Y} \text{ in a single point,}\\ &= \text{undefined, otherwise.}\end{aligned}$$

The map $\pi_{\mathbf{O}}$ is called the *central projection with center* $\boldsymbol{O}$ *of* $\boldsymbol{R}^n$ *to the plane* $\boldsymbol{Y}$. If $\mathbf{X}$ is another hyperplane in $\mathbf{R}^n$, then the restriction of $\pi_{\mathbf{O}}$ to $\mathbf{X}$, $\pi_{\mathbf{O}}|\mathbf{X} : \mathbf{X} \to \mathbf{Y}$, is called the *perspective transformation* or *perspectivity from* $\boldsymbol{X}$ *to* $\boldsymbol{Y}$ *with center* $\boldsymbol{O}$.

Note that our terminology makes a slight distinction between central projections and perspectivities. Both send points to a plane, but the former is defined on all of Euclidean space, whereas the latter is only defined on a plane; however, they clearly are closely related.

Clearly, from the point of view of formulas, one would not expect our new maps to be complicated because they simply involve finding the intersection of a line with a hyperplane. Let us look at some simple examples to get a feel for what geometric properties these maps possess. First, consider perspectivities between lines in $\mathbf{R}^2$. Figure 3.1 shows the case where the two lines **parallel**. In this case, the ratio of the distance between points and the distance between their images is constant. The perspectivity is one-to-one and onto. It preserves parallelism, concurrence, ratio of division, and betweenness.

What happens when the two lines are **not** parallel? See Figure 3.2. The point $\mathbf{V}$ on $\mathbf{L}$ has no image and the point $\mathbf{W}$ on $\mathbf{L}'$ has no preimage. These points are called *vanishing points*. Betweenness is **not** preserved as is demonstrated by the points $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in Figure 3.2. Furthermore, the fact that betweenness is not preserved leads to other properties not being preserved. In particular, segments, rays, and ratios of division are not preserved, and distances are distorted by different constants.

Next, consider perspectivities between planes. When the planes are parallel, things behave pretty well just like for parallel lines. The interesting case is when the planes are not parallel. Consider a perspectivity with center $\mathbf{O}$ from a plane $\mathbf{X}$, which we shall call the object plane, to another plane $\mathbf{Y}$, which we shall call the view plane. The following facts are noteworthy.

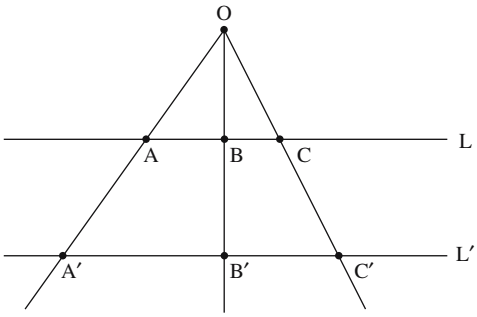



**Figure 3.1.** A perspectivity between parallel lines.

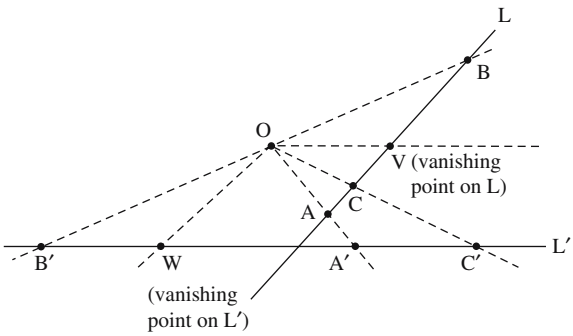



**Figure 3.2.** A perspectivity between skew lines.

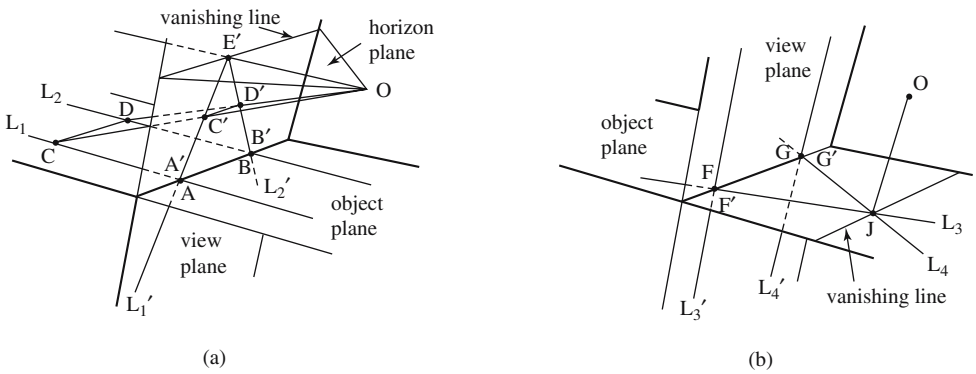



**Figure 3.3.** A perspectivity between skew planes.

See Figure 3.3(a).

(1) The rectangle **ABCD** in the object plane gets mapped to the trapezoid **A′B′C′D′** and the lines $\mathbf{L}_1$ and $\mathbf{L}_2$ through **A**,**C** and **B**,**D**, respectively, get mapped to the lines $\mathbf{L}_1'$ and $\mathbf{L}_2'$ through **A′**,**C′** and **B′**,**D′**.

(2) The parallel lines $\mathbf{L}_1$ and $\mathbf{L}_2$ get mapped to lines that intersect at a point **E′** in the view plane. The point **E′** has the property that **OE′** is parallel to the object plane.

The point **E′** is called a *vanishing point*. The *horizon plane*, which is the plane through **OE′** parallel to the object plane, intersects the view plane in a line called a *vanishing line* or *horizon*. Note that the perspectivity is undefined at every point of the vanishing line.

(3) Not only **E′** but **each** point of the vanishing line is called a *vanishing point*. One can show that every such vanishing point comes from the projection of a family of parallel lines. Another way to put this is that a perspectivity between nonparallel planes maps parallel lines into intersecting lines.

For the next facts, see Figure 3.3(b).

(4) The intersecting lines $\mathbf{L}_3$ and $\mathbf{L}_4$ through **F**,**J** and **G**,**J**, respectively, map to parallel lines $\mathbf{L}_3'$ and $\mathbf{L}_4'$ in the view plane. The point **J** has the property that **OJ** is parallel to the view plane. The point **J** is also called a *vanishing point*. The plane through **OJ** parallel to the view plane, intersects the object plane in a line also called a *vanishing line*. Not only **J** but **each** point of this vanishing line is called a *vanishing point*. Note that the perspectivity maps no point of $\mathbf{R}^n$ onto any point of the vanishing line in the object plane.

(5) One can show that every collection of lines in the object plane that intersect in a point on the vanishing line map to parallel lines in the view plane. Again, another way to put this is that a perspectivity between nonparallel planes maps every family of lines that intersect on the vanishing line in the object plane into a family of parallel lines.

(6) Closed figures can go to open figures and vice versa. For example, the triangle **FGJ** gets mapped to an unbounded region in the view plane bounded by the segment [**F**,**G**] and the lines $\mathbf{L}'_3$ and $\mathbf{L}'_4$.

Another fact that is true but not explicitly shown in Figure 3.3 is:

(7) Circles can project onto ellipses, parabolas onto hyperbolas; in fact, every nondegenerate conic can project into any other nondegenerate conic (see Section 3.6.1 and Theorem 3.6.1.1).

As an interesting aside, a consequence of the above is that the reason that we see a "horizon" when looking out over the ocean has nothing to do with the fact that the earth is round. We would see this even if the earth were flat. The only difference is that, in our case of a round earth, one sees the mast of an approaching ship on the horizon before one sees its hull. This would not happen if the earth were flat.

Perspectivities are a special case of what are called *projective transformations*. A quick definition of these would be to say that they are composites of perspectivities, but, keeping in line with the way we developed the affine transformations, we shall make this a theorem and phrase our definition in terms of invariant properties. Furthermore, in this section the goal will be to outline the theory of projective transformations **within** affine space ($\mathbf{R}^n$). Let us see how far we can go.

**Definition.** Properties that are preserved by parallel and central projections are called *projective properties* or *projective invariants*.

The figures above pointed out aspects of some specific perspectivities. What are the abstract invariant properties that characterize such maps in general? Earlier we mentioned some geometric properties that perspectivities between lines do not preserve. Is there anything that all perspectivities preserve? Yes, there is, and it is called the cross-ratio.

**Definition.** Let **A**, **B**, **C**, and **D** be distinct collinear points on an **oriented** line **L**. The *cross-ratio* in which **B** and **C** divide **A** and **D**, denoted by (**AD**,**BC**), is defined to be the following quotient of ratios of division:

$$(\mathbf{AD},\mathbf{BC}) = \frac{(\mathbf{AD},\mathbf{B})}{(\mathbf{AD},\mathbf{C})} = \frac{\|\mathbf{AB}\|}{\|\mathbf{BD}\|}\bigg/\frac{\|\mathbf{AC}\|}{\|\mathbf{CD}\|} = \frac{\|\mathbf{AB}\|}{\|\mathbf{BD}\|}\cdot\frac{\|\mathbf{CD}\|}{\|\mathbf{AC}\|},$$

where $\|\ \|$ denotes the **signed** distance between points of the **oriented** line **L**.

Although developed by the ancient Greeks, the modern development of the cross-ratio is due to A.F. Moebius (*Der Barycentrische Calcul*, 1827) and independently to M. Chasles (various publications from 1829–1865). The term "cross-ratio" was coined by W.K. Clifford in 1878.

Because the ratios of division are independent of the orientation of the line, so is the cross-ratio. The cross-ratio can be (and is often) defined for points **A**, **B**, **C**, and **D**, where only **three** of those four points are distinct if one defines it to be ∞ in the duplicate point case. We shall not do so here.

To explain the somewhat mysterious concept of cross-ratio, we first look at the case of four numbers a, b, c, and d. See Figure 3.4. By definition

$$(\mathrm{ad},\mathrm{bc}) = \frac{\mathrm{b}-\mathrm{a}}{\mathrm{d}-\mathrm{b}}\bigg/\frac{\mathrm{c}-\mathrm{a}}{\mathrm{d}-\mathrm{c}} = \frac{\mathrm{b}-\mathrm{a}}{\mathrm{d}-\mathrm{b}}\frac{\mathrm{d}-\mathrm{c}}{\mathrm{c}-\mathrm{a}}. \tag{3.1}$$

Figure 3.4(a) shows the intervals involved in the formula. Figure 3.4(b) shows some values of (ad,bc), where we fixed a, c, and d and let b vary. We see that the cross-ratio changes from (d – c)/(c – a) to 0 as b increases from –∞ to a. It then increases from 0 to ∞ as b increases from a to d. The cross-ratio decreases from (d – c)/(c – a) to –∞ as b decreases from +∞ to d. In general, if one fixes three distinct points **A**, **C**, and **D** on a line **L**, then the function

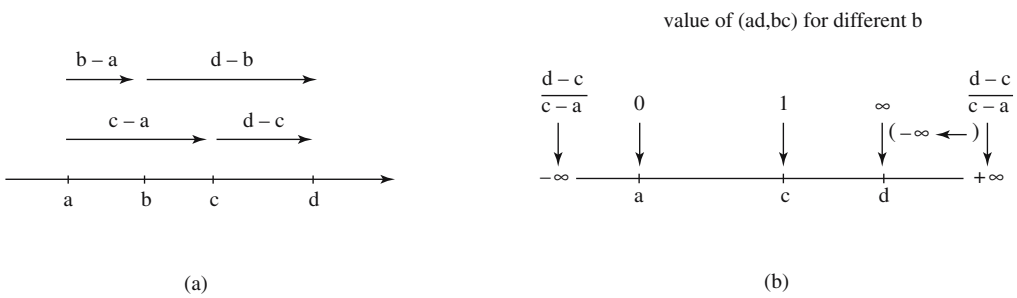



**Figure 3.4.** The cross-ratio for four numbers.

$$(\mathbf{AD},-\mathbf{C}):\ \mathbf{L}-\{\mathbf{D}\}\to\mathbf{R}$$
$$\mathbf{B}\ \to(\mathbf{AD},\mathbf{BC})$$

is a one-to-one, onto, and continuous function. This fact is easily deduced from equation (3.1).

Now, it is easy to check that the formula on the right-hand side of equation (3.1) is invariant under affine maps. Although the x-axis may seem like a very special line, all lines look like it once they are parameterized. We prove the following fundamental theorem of projective geometry.

**3.2.1. Theorem.** Both parallel projections and perspectivities between hyperplanes preserve the cross-ratio.

**Proof.** By Theorem 2.4.1.2 parallel projections preserve the ratio of division. Therefore, parallel projections preserve the cross-ratio. Before we get to a proof of the theorem in the case of perspectivities, we derive a formula for the ratio of division. Assume that we have points **A**, **B**, and **D** on an oriented line **L** and that **O** is a point not on the line. See Figure 3.5(a). If $\mathbf{n_A}$ and $\mathbf{n_D}$ are unit vectors that are orthogonal to the vectors **OA** and **OD**, respectively, then

$$\mathbf{AB}\bullet\mathbf{n_A}=\mathbf{OB}\bullet\mathbf{n_A}, \tag{3.2}$$

and

$$\mathbf{BD}\bullet\mathbf{n_D}=\mathbf{BO}\bullet\mathbf{n_D}, \tag{3.3}$$

because the quantities in equations (3.2) and (3.3) are just the signed lengths of the perpendiculars dropped from **B** to the lines through **O**, **A**, and **O**, **D**, respectively.

Next, let **u** be a **unit** direction vector for **L** that induces its orientation. Define real numbers a and b by

$$\mathbf{AB}=a\mathbf{u}, \tag{3.4}$$

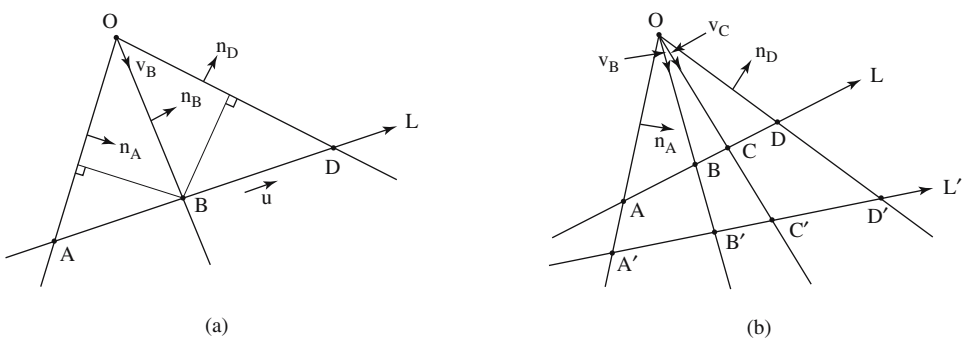



**Figure 3.5.** Perspectivities preserve the cross-ratio.

and

$$\mathbf{BD} = b\mathbf{u}. \tag{3.5}$$

Note that a and b are just the signed distances between **A**, **B**, and **B**, **D**, respectively. Equations (3.4) and (3.5) imply that

$$\mathbf{AB} \bullet \mathbf{n_A} = a\mathbf{u} \bullet \mathbf{n_A} \tag{3.6}$$

and

$$\mathbf{BD} \bullet \mathbf{n_D} = b\mathbf{u} \bullet \mathbf{n_D}. \tag{3.7}$$

Finally, let $\mathbf{v_B}$ be the unit vector in direction **OB**, that is,

$$\mathbf{OB} = |\mathbf{OB}|\mathbf{v_B}. \tag{3.8}$$

Equations (3.2), (3.3), (3.6), (3.7), and (3.8) lead to the following formula for the ratio of division (**AD**,**B**):

$$(\mathbf{AD},\mathbf{B}) = \frac{a}{b} = \frac{(\mathbf{AB} \bullet \mathbf{n_A})(\mathbf{u} \bullet \mathbf{n_D})}{(\mathbf{BD} \bullet \mathbf{n_D})(\mathbf{u} \bullet \mathbf{n_A})} = \frac{\mathbf{OB} \bullet \mathbf{n_A}}{\mathbf{BO} \bullet \mathbf{n_D}} \frac{\mathbf{u} \bullet \mathbf{n_D}}{\mathbf{u} \bullet \mathbf{n_A}} = -\frac{\mathbf{v_B} \bullet \mathbf{n_A}}{\mathbf{v_B} \bullet \mathbf{n_D}} \frac{\mathbf{u} \bullet \mathbf{n_D}}{\mathbf{u} \bullet \mathbf{n_A}}. \tag{3.9}$$

This formula shows why the ratio of division is not a projective invariant. It depends on the direction vector **u** of the line **L** containing the points.

We return to the proof of Theorem 3.2.1. Consider Figure 3.5(b). We need to show that (**AD**,**BC**) = (**A′D′**,**B′C′**). If we define

$$\mathbf{AC} = c\mathbf{u}, \quad \mathbf{CD} = d\mathbf{u}, \quad \text{and} \quad \mathbf{OC} = |\mathbf{OC}|\mathbf{v_C},$$

then a derivation similar to the one that led to equation (3.9) shows that

$$(\mathbf{AD},\mathbf{C}) = \frac{c}{d} = -\frac{\mathbf{v_C} \bullet \mathbf{n_A}}{\mathbf{v_C} \bullet \mathbf{n_D}} \frac{\mathbf{u} \bullet \mathbf{n_D}}{\mathbf{u} \bullet \mathbf{n_A}}. \tag{3.10}$$

Equations (3.9) and (3.10) show that

$$(\mathbf{AD},\mathbf{BC}) = \frac{(\mathbf{AD},\mathbf{B})}{(\mathbf{AD},\mathbf{C})} = \frac{\mathbf{v_B} \bullet \mathbf{n_A}}{\mathbf{v_B} \bullet \mathbf{n_D}} \frac{\mathbf{v_C} \bullet \mathbf{n_D}}{\mathbf{v_C} \bullet \mathbf{n_A}}. \tag{3.11}$$

The right-hand side of equation (3.11) depends only on angles defined by the rays through **O** (actually the sines of the angles between the rays) and not on where the points lie on the rays. Dividing the two ratios of division eliminated the other dependencies. Therefore, we would get the same value for (**A′D′**,**B′C′**), specifically because $\mathbf{v_B} = \mathbf{v_B}'$ and $\mathbf{v_C} = \mathbf{v_C}'$. Theorem 3.2.1 is proved.

There is an important and computationally useful consequence of the projective invariance of the cross-ratio. See Figure 3.6. If $\mathbf{A}(x_1,y_1)$, $\mathbf{B}(x_2,y_2)$, $\mathbf{C}(x_3,y_3)$, and $\mathbf{D}(x_4,y_4)$

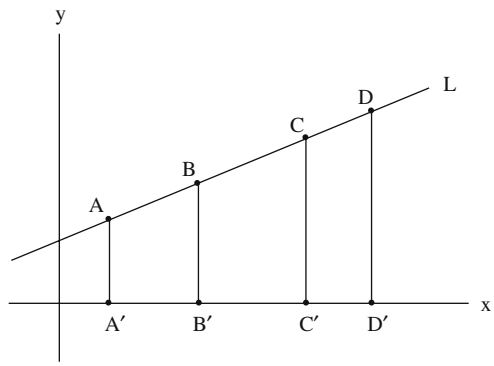



**Figure 3.6.** Computing the cross-ratios from x-coordinates.

are distinct collinear points that have distinct x-coordinates, then one can compute their cross-ratio from their x-coordinate values. In other words, if $\mathbf{A}'(x_1,0)$, $\mathbf{B}'(x_2,0)$, $\mathbf{C}'(x_3,0)$, and $\mathbf{D}'(x_4,0)$ are the projections of the points on the x-axis, then

$$(\mathbf{AD},\mathbf{BC}) = (\mathbf{A'D'},\mathbf{B'C'}) = \frac{x_2 - x_1}{x_4 - x_2} \Big/ \frac{x_3 - x_1}{x_4 - x_3} = \frac{x_2 - x_1}{x_4 - x_2}\,\frac{x_4 - x_3}{x_3 - x_1}.$$

Similarly, one can use the y-coordinates if those are distinct.

**3.2.2. Example.** To compute the cross-ratio of the points $\mathbf{A}(2,0)$, $\mathbf{B}(0,-1)$, $\mathbf{C}(6,2)$, and $\mathbf{D}(8,3)$ on the line $\mathbf{L}$ defined by the equation $x - 2y - 2 = 0$, assuming that $\mathbf{L}$ is oriented to the right.

**Solution.** By definition of the cross-ratio, we have that

$$(\mathbf{AD},\mathbf{BC}) = \frac{\|\mathbf{AB}\|}{\|\mathbf{BD}\|}\cdot\frac{\|\mathbf{CD}\|}{\|\mathbf{AC}\|} = \frac{\sqrt{5}}{2\sqrt{5}}\cdot\frac{-\sqrt{5}}{4\sqrt{5}} = -\frac{1}{8}.$$

Since the x-coordinates 2, 0, 6, and 8 of the points are distinct, an easier way to compute the cross-ratio is to use these values and formula (3.1):

$$\frac{0-2}{8-0}\cdot\frac{8-6}{6-2} = -\frac{1}{8}$$

The answer is the same.

We point out another interesting geometric consequence of Theorem 3.2.1. Since the cross-ratio is preserved by a perspectivity, it follows that the view of four collinear points is completely determined once one knows the position of three of those points in the view. For example, consider a railroad track that consists of two rails and equally spaced ties or a ladder with equally spaced rungs. What are the possible perspective views of this track or ladder? Well, the position and relative spacing of any three of these ties or rungs in the view can be quite arbitrary except for some minor

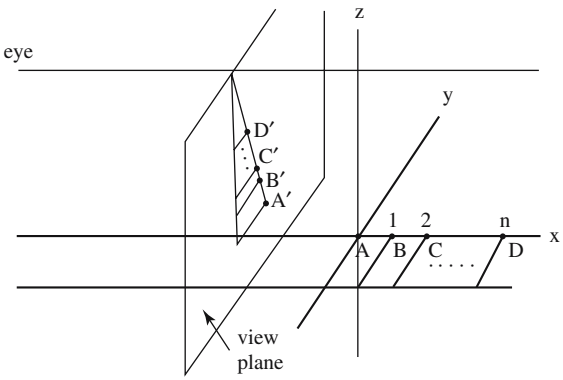



**Figure 3.7.** The view of three railroad ties determines the rest.

constraints, but from then on the relative spacing of the remaining ties or rungs in the view is fixed. See Figure 3.7.

Returning to our analysis of parallel and central projections, the next definition is the best we can do in the context of affine geometry.

**Definition.** A *projective transformation* or *projectivity* of a plane **X** to itself is a "map" T such that

(a) The domain of T is either **X** or **X** minus a line **L**. The range of T is either **X** or **X** minus a line **L**′.
(b) T is one-to-one.
(c) T preserves collinearity when defined.
(d) T preserves the cross-ratio when defined.

**3.2.3. Theorem.** A projective transformation has the following properties:

(1) Ordinary lines go to ordinary lines.
(2) The map sets up a natural one-to-one correspondence between lines.
(3) A family of concurrent lines goes into a family of concurrent lines or a family of parallel lines.
(4) A family of parallel lines goes into a family of parallel lines or a family of concurrent lines.
(5) Conics go to conics. More precisely, any nondegenerate conic may be mapped onto any other nondegenerate conic by a projective transformation.

**Proof.** See [Gans69]. Theorem 3.2.3 can be proved directly or deduced from Theorem 3.2.5 below.

**3.2.4. Theorem.** There is a unique projective transformation in which a given line **L** has no image and which sends given noncollinear points **A**, **B**, and **C** into given noncollinear points **A**′, **B**′, and **C**′, respectively.

**Proof.** See [Gans69].

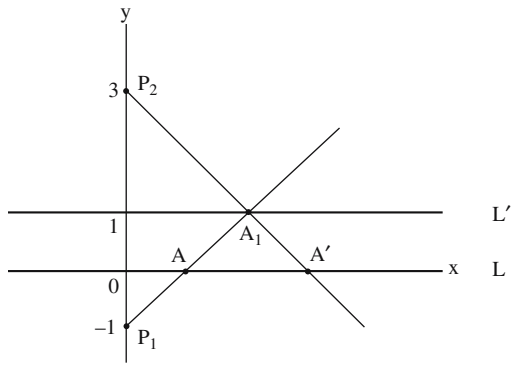



**Figure 3.8.** A similarity produced by two central projections.

The next theorem gives a nice geometric characterization of projective transformations and generalizes Theorem 2.4.1.2.

**3.2.5. Theorem.** Every projective transformation is the composite of a sequence of parallel or central projections not more than one of which is central. Conversely, every transformation of the plane that is a composite of parallel and central projections is a projective transformation.

**Proof.** See [Gans69].

In other words, projective transformations are the transformations that preserve projective properties.

**3.2.6. Example.** The similarity T: $\mathbf{R} \to \mathbf{R}$ defined by $T(x) = 3x$ is a composite of two central projections with centers at $\mathbf{P}_1 = (0,-1)$ and $\mathbf{P}_2 = (0,3)$. See Figure 3.8.

**3.2.7. Theorem.** (The Fundamental Theorem of Projective Geometry for the Plane) There is a unique projective transformation that sends four given points **A**, **B**, **C**, and **D** into another four given points **A′**, **B′**, **C′**, and **D′**, respectively, if no three in either set are collinear.

**Proof.** See [Gans69].

Finally, there is the question as to what the equations for a projective transformation might look like.

**Definition.** A transformation of the plane of the form

$$x' = \frac{a_1x + a_2y + a_3}{c_1x + c_2y + c_3}, \quad y' = \frac{b_1x + b_2y + b_3}{c_1x + c_2y + c_3},$$

where

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0$$

is called a *fractional transformation*.

**3.2.8. Theorem.** Every projective transformation is a fractional transformation and conversely.

**Proof.** See [Gans69].

**3.2.9. Theorem.** The set of all projective transformations is a group containing the planar affine transformations as a subgroup.

**Proof.** This is obvious. The planar affine transformations correspond to the fractional transformations above where $c_1 = c_2 = 0$ and $c_3 = 1$.

This is as far as we shall take things in affine space. Clearly, one of the unpleasant aspects of perspectivities in this context is that they are not defined everywhere, nor are they onto. Basically, there are "missing" points. We shall have more to say about that shortly, but a few comments are appropriate now because it prepares the reader for the concept of "ideal point" and we also want to relate perspectivities to parallel projections. We know what the missing points are in the case of a perspectivity of the type shown in Figure 3.3. We need to add some points, both to the view plane and the object plane. Suppose we add a new point to a plane for each of its families of parallel lines. If **L** is a line, then let $\mathbf{L}^\infty$ denote this new point associated to the family of lines parallel to **L**. $\mathbf{L}^\infty$ will be called an *ideal point*. With these new points, we could extend our definition of the perspectivity by saying that the point **J** in Figure 3.3(b) should map to $\mathbf{L}_3^\infty$ and $\mathbf{L}_1^\infty$ should map to $\mathbf{E}'$ in Figure 3.3(a). This would give us a one-to-one and onto map between these *extended planes*, which are ordinary planes together with their ideal points. In addition, with these new points we could consider **parallel projections between hyperplanes as a special case of perspectivities** if we allow the center of projection to be an ideal point.

## 3.3 Homogeneous Coordinates

One of the key ideas in the study of analytic projective geometry is that of **homogeneous coordinates**. The standard Cartesian coordinates are sometimes referred to as "nonhomogeneous" coordinates and are simply **one** of **many** ways to specify points in space with real numbers. Other ways are polar coordinates in the plane and cylindrical and spherical coordinates in 3-space. Barycentric coordinates are a type of "homogeneous" coordinates. They specify points relative to a fixed set of points.

Out of the many ways that one can coordinatize points, which is the most convenient depends completely on the type of problem we are trying to solve. Homogeneous coordinates are just another way of coordinatizing points. Historically they find

their roots in Moebius' work on barycentric coordinates (*Der Barycentrische Calcul*, 1827) and the fact that they are useful with central projections. Here we motivate their definition by looking at the relationship between points and solutions to linear equations.

We shall start with the real line **R**. What we are about to do may seem a little silly at first, but if the reader will bear with us, it should make more sense in the end. Linear equations in **R** have the form

$$ax + b = 0, \quad \text{with} \quad a \neq 0. \tag{3.12}$$

We can think of **R** as the set of solutions to all equations of the form (3.12). Equation (3.12) is homogeneous in a and b, but not in x. We can achieve more symmetry by introducing another variable Y and consider the equation

$$aX + bY = 0, \quad \text{with} \quad (a,b) \neq (0,0). \tag{3.13}$$

The trivial solution (X,Y) = (0,0) is uninteresting and will be **excluded** from consideration. Note that if we have a solution x to equation (3.12), then we have a solution (x,1) to equation (3.13). In fact, (kx,k) will also be a solution to (3.13) for all $k \neq 0$. Conversely, if (X,Y) is a solution to (3.13), then X/Y is a solution to (3.12) **if** $Y \neq 0$. In short, each solution x to (3.12) gives rise to a class of solutions (kx,k), $k \neq 0$, to (3.13) and each class of solutions (kX,kY) to (3.13) with $k \neq 0$ and $Y \neq 0$ gives rise to a unique solution X/Y to (3.12).

**Definition.** Let $x \in \mathbf{R}$ and let (X,Y) be any pair of real numbers with $Y \neq 0$. If x = X/Y, then X and Y are called *homogeneous coordinates* for x. One typically uses the expression "(X,Y) are homogeneous coordinates for x" in that case.

Note that if (X,Y) are homogeneous coordinates for a real number x, then so are (kX,kY) for any $k \neq 0$.

**3.3.1. Example.** (–2,1), (–4,2), and (8,–4) are all homogeneous coordinates for the real number –2.

What about the solutions to (3.13) with Y = 0? Since $X \neq 0$ in that case, they are all multiples of (1,0). In conclusion, all the solutions to (3.13) fall into classes each of which corresponds to a unique solution to (3.12) except for the one extra class of solutions (k,0). We can think of these classes as **points**. The set of these "points" will be called the "projective line." It can be thought of as consisting of the real numbers with one additional point added (the one corresponding to Y = 0). The extra point is called an *ideal point*. All this will be formalized in the next section.

Next, we look at the more interesting case of the plane. The essential ingredients in the analytic development of the affine (Cartesian) plane are points, which are pairs (x,y), and lines, which are the set of solutions (x,y) to an equation of the form

$$ax + by + c = 0, \quad \text{with} \quad (a,b) \neq (0,0). \tag{3.14}$$

Note the following semi-duality in this context:

(a) Points on a line are the solutions (x,y) to equation (3.14) keeping a, b, and c fixed.
(b) Lines (through the fixed point (x,y)) are the solutions (a,b,c) to equation (3.14) keeping (x,y) fixed.

In (a) we are dealing with nonhomogeneous solutions and in (b) they are homogeneous. We can make the solutions homogeneous in both cases by introducing a new variable Z. Furthermore, to obtain complete symmetry, we restate the equation (3.14) in the form

$$aX + bY + cZ = 0, \quad \text{with} \quad (a,b,c) \neq (0,0,0). \tag{3.15}$$

Because equation (3.15) always has the trivial solution (0,0,0), and because this solution is uninteresting, we shall always **exclude** it from any discussion.

Clearly, if (X,Y,Z) is a solution to (3.15) with $Z \neq 0$, then (X/Z,Y/Z) is a solution to (3.14). Conversely, if (x,y) is a solutions to (3.14), then (x,y,1) is a solution to (3.15), as is (kx,ky,k) for any real k. What this discussion is leading up to is that instead of thinking of points as coordinatized by pairs (x,y) of real numbers we can think of them as coordinatized by triples (X,Y,Z), or rather by classes of triples.

**Definition.** Let **p** be any point in the plane $\mathbf{R}^2$. Let (X,Y,Z) be any triple of real numbers with $Z \neq 0$. If x = X/Z and y = Y/Z are the Cartesian coordinates for **p**, then X, Y, and Z are called *homogeneous coordinates* for **p**. One typically uses the expression "(X,Y,Z) are homogeneous coordinates for **p**" in that case.

Note that if (X,Y,Z) are homogeneous coordinates for a point (x,y), then so are (kX,kY,kZ) for any $k \neq 0$.

**3.3.2. Example.** (–1,2,1), (2,–4,–2), and (–4,8,4) are all homogeneous coordinates for the Cartesian point (–1,2).

So far in our correspondence between the solutions to (3.14) and (3.15) we excluded the solutions to (3.15) that have Z = 0. What about these solutions?

Suppose that (X,Y,0) is a solution to (3.15) and that $(a,b) \neq (0,0)$.

**Case 1.** $a \neq 0$: If Y = 0, then X = 0. Since the solution (0,0,0) is not allowed, we must have $Y \neq 0$ and X = (–b/a)Y. In other words, every solution has the form ((–b/a)Y,Y,0).

**Case 2.** $b \neq 0$: This time we can solve for Y and every solution has the form (X,(–a/b)X,0), with $X \neq 0$.

We can combine cases 1 and 2 to conclude that every solution must have the form (–bt,at,0) for some $t \neq 0$. Alternatively, we could have deduced this from the observation that the only vectors (X,Y) that are orthogonal to (a,b) are scalar multiples of (–b,a). In any case, we see that there is only one extra class of solutions. Let us call each class of solutions (X,Y,Z) to (3.15) informally a **point** and the set of these "points." the "projective plane." (Precise definitions are given in the next section.). The extra points where Z = 0 will be called *ideal* points.

Recall that equations for lines are not unique. Also, if **L** is the line defined by equation (3.14), then the family of lines parallel to **L** is obtained as solutions to equations where we fix a and b in (3.14) but let c vary (Exercise 1.5.7(b)). Looking at this another way, what we have just shown is that there is a one-to-one correspondence between families of parallel lines in the plane and ideal points, namely, to the family of lines parallel to **L** we associate the ideal point that is the class of solutions to (3.15) determined by (–b,a,0). That such a correspondence exists was foreshadowed in the discussion at the end of Section 3.2.

Finally, the equations in (3.14) give rise to all of the equations in (3.15) **except** the equation $cZ = 0$, that is, $Z = 0$. Thus there is only one equation in (3.15) not arising from a line in the plane, but this is precisely the equation that defines the ideal points. It should not be surprising if, as we shall do in the next section, one defines a "line" in the projective plane to mean a set of points determined by the solutions to a linear equation of the form shown in (3.15). Then lines in the projective plane will correspond to solutions to linear equations just like in the Euclidean plane.

## 3.4 The Projective Plane

The informal discussion of linear equations and their solutions in the previous section led to homogeneous coordinates and suggested a new way of looking at points in the plane. We shall now develop these observations more rigorously. Although we are only interested in the projective line and plane for a while, we start off with some general definitions so that we do not have to repeat them for each dimension.

**3.4.1. Lemma.** The relation ~ defined on the points **p** of $\mathbf{R}^{n+1} - \mathbf{0}$ by $\mathbf{p} \sim c\mathbf{p}$, for $c \neq 0$, is an equivalence relation.

**Proof.** This is an easy exercise.

**Definition.** The set of equivalence classes of $\mathbf{R}^{n+1} - \mathbf{0}$ with respect to the relation ~ defined in Lemma 3.4.1 is called the *n-dimensional (real) projective space* $\mathbf{P}^n$. In more compact notation (see Section 5.4 and the definition of a quotient space),

$$\mathbf{P}^n = \left(\mathbf{R}^{n+1} - \mathbf{0}\right)/\sim.$$

The special cases $\mathbf{P}^1$ and $\mathbf{P}^2$ are called the *projective line* and *projective plane*, respectively. If $\mathbf{P} \in \mathbf{P}^n$ and $\mathbf{P} = [x_1, x_2, \ldots, x_{n+1}]$, then the numbers $x_1, x_2, \ldots, x_{n+1}$ are called *homogeneous coordinates* of **P**. One again typically uses the expression "$(x_1, x_2, \ldots, x_{n+1})$ are homogeneous coordinates for **P**" in that case.

Note that $\mathbf{P}^0$ consists of the single point [1]. We can think of points in $\mathbf{P}^1$ or $\mathbf{P}^2$ as equivalence classes of solutions to (3.13) or (3.15), or alternatively, as the set of lines through the origin in $\mathbf{R}^2$ or $\mathbf{R}^3$, respectively. Other characterizations of the abstract spaces $\mathbf{P}^n$ will be given in Section 5.9. There are actually many ways to introduce coordinates for their points. In the next section we shall see how this can be done for $\mathbf{P}^1$ and $\mathbf{P}^2$.

It is easy to check that the maps

$$\begin{array}{ccccccc} \mathbf{P}^0 & \to & \mathbf{P}^1 & , & \mathbf{P}^1 & \to & \mathbf{P}^2 \\ [X] & \to & [0,X] & & [X,Y] & \to & [X,0,Y] \end{array}$$

and

$$\begin{array}{ccccccc} \iota : \mathbf{R} & \to & \mathbf{P}^1 & , & \iota : \mathbf{R}^2 & \to & \mathbf{P}^2 \\ x & \to & [x,1] & & (x,y) & \to & [x,y,1] \end{array}$$

are one-to-one. Therefore, by identifying the corresponding points, we shall think of these maps as inclusion maps and get a commutative diagram

$$\begin{array}{ccccc} \mathbf{P}^0 & \subset & \mathbf{P}^1 & \subset & \mathbf{P}^2 \\ \| & & \cup & & \cup \\ \mathbf{R}^0 & \subset & \mathbf{R}^1 & \subset & \mathbf{R}^2. \end{array}$$

The "inclusion" maps $\iota$ above are called the *standard imbeddings* of $\mathbf{R}$ in $\mathbf{P}^1$ and $\mathbf{R}^2$ in $\mathbf{P}^2$. In particular, we shall consider the Euclidean plane as a **subset** of the projective plane.

The two notions of "homogeneous coordinates" and "projective space" are really quite inseparable. Homogeneous coordinates define the natural coordinate systems for projective space. Even if they are only used to coordinatize points of Euclidean space, one is dealing with projective space at least implicitly.

**Definition.** The points of $\mathbf{P}^n - \mathbf{R}^n$ are called *ideal points*. All the other points are called *real points*.

The projective line $\mathbf{P}^1$ has only one ideal point [1,0], which we shall denote by $\infty$ and so, as a set, $\mathbf{P}^1$ can be identified with the union $\mathbf{R} \cup \{\infty\}$. Furthermore, with the natural topology, the real points of the space that "converge" to $\pm\infty$ as numbers would converge to the ideal point $\infty$. In other words, $\mathbf{P}^1$ is a circle topologically.

The projective plane $\mathbf{P}^2$ is another space that is probably new to the reader. What does it really "look" like topologically? We shall postpone a careful answer to this question to Chapter 5 because this chapter has different goals. However, note that the ideal points of $\mathbf{P}^2$ are the points of the form [X,Y,0]. There is a natural correspondence between those points and the points [X,Y] of $\mathbf{P}^1$, that is, the set of ideal points in the projective plane look like a copy of the projective line. To put it another way, a good way to think of the projective plane is as the standard Euclidean plane with a circle added at infinity. As sets, $\mathbf{P}^2 = \mathbf{R}^2 \cup \mathbf{P}^1$. In Section 5.9 we shall give some other definitions that produce spaces topologically equivalent to the space $\mathbf{P}^2$ defined above. For example, we shall see that we can think of the projective plane as a disk with antipodal points on its boundary identified. The points derived from the boundary correspond to the ideal points. See Figure 3.9(a). This also leads to another way of thinking of $\mathbf{P}^2$, namely, as the union of a Moebius strip and a disk with the two boundaries (both are circles) glued together. See Figure 3.9(b). In this chapter, however, we are not interested in the projective plane from the point of view of topology but in terms of its algebraic and analytic structure. As an analogy, note that $\mathbf{R}^2$ looks like an

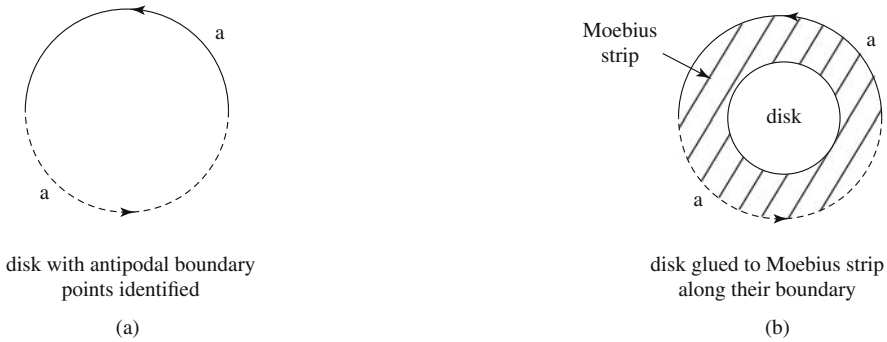



**Figure 3.9.** Visualizing the projective plane.

open disk topologically, but that is not a useful way of looking at it if one wants to use its vector space structure.

It should be pointed out that there is actually nothing special about ideal points even though their name would suggest otherwise. From an intrinsic point of view, every point of $\mathbf{P}^n$ looks like every other point of $\mathbf{P}^n$. The reader needs to understand what really happened here. Projective spaces are **abstract** sets (sets of equivalence classes) and we decided to coordinatize their points, that is, we decided to associate a tuple of numbers with each point. Although our choice of coordinates is a natural one, it is only one out of infinitely many ways that one can assign coordinates to the points of these abstract sets. In the case of $\mathbf{P}^2$, our chosen coordinates imbedded $\mathbf{R}^2$ in $\mathbf{P}^2$ in a special way. Which points end up being called ideal points is totally a function of how $\mathbf{R}^2$ is imbedded. Their special nature is purely an artifact of the construction. For example, one can think of a sphere as the Euclidean plane with one extra point "at infinity" added. This would seem to make this one point special, but it is only special because of the particular representation. A sphere, thought of in the abstract, has all points and their neighborhoods looking the same. We shall return to the issue of coordinates in the next section and Sections 8.13 and 10.3.

**Notation.** If $\ell$ is the line in the plane defined by equation (3.14), let $\ell^\infty$ denote the ideal point [–b,a,0].

It follows from the discussion in the previous section that the association of $\ell^\infty$ with the family of lines parallel to $\ell$ sets up a one-to-one correspondence between ideal points and families of parallel lines in the plane.

**Definition.** A *line* in $\mathbf{P}^2$ is any subset $\mathbf{L}$ of $\mathbf{P}^2$ of the form

$$\mathbf{L} = \{[X,Y,Z] \mid aX + bY + cZ = 0 \quad \text{for fixed} \quad (a,b,c) \neq (0,0,0)\}.$$

Points of $\mathbf{P}^2$ that lie on a line are said to be *collinear*.

The tuple (a,b,c) that defines the line $\mathbf{L}$ above is not unique because any nonzero multiple defines the same line. On the other hand, it is easy to show that the

correspondence between **L** and [a,b,c] sets up a one-to-one and onto correspondence between lines **L** in $\mathbf{P}^2$ and points [a,b,c] in $\mathbf{P}^2$.

**Notation.** Given a line $\ell$ in $\mathbf{R}^2$ defined by the equation

$$ax + by + c = 0$$

or a line **L** in $\mathbf{P}^2$ defined by the equation

$$aX + bY + cZ = 0,$$

we shall let $[\ell]$ and [**L**] denote the point [a,b,c] in $\mathbf{P}^2$.

With this notation, consider a line **L** and assume that [**L**] = [**a**], where $\mathbf{a} \in \mathbf{R}^3$. It follows that **L** is just the set of all points [**p**] in $\mathbf{P}^2$, so that

$$\mathbf{a} \bullet \mathbf{p} = 0. \tag{3.16}$$

This gives us a complete duality between points on a line and lines on a point in $\mathbf{P}^2$. This observation was already implicit in our discussion earlier of equation (3.15). We make this duality a little more explicit.

**Definition.** Given a statement concerning the projective plane, the *dual* of that statement is the statement where every occurrence of the words "point" or "collinear" is replaced by "line" or "concurrent," respectively, and vice versa.

**The principle of duality in the projective plane:** Given any theorem for the projective plane, its dual is also a theorem.

The duality principle is very useful in the study of projective geometry. Even though it will only get an occasional mention in this book, it is worth being aware of it.

All lines are alike in $\mathbf{P}^2$ (topologically they are circles), but with our imbedding of $\mathbf{R}^2$ in $\mathbf{P}^2$, we can distinguish between two types of lines.

**Definition.** An *ordinary* line **L** in $\mathbf{P}^2$ is a line that consists of an ordinary line $\ell$ in the Euclidean plane with an ideal point adjoined, that is, a line of the form $\mathbf{L} = \ell \cup \{\ell^\infty\}$. (See Exercise 3.4.1.) The line defined by the equation Z = 0, which consists of the set of ideal points, is called the *ideal* line.

Note that there is no such thing as "parallel" lines in the projective plane.

**3.4.2. Theorem.** Every pair of distinct lines in $\mathbf{P}^2$ intersects in a point.

**Proof.** Since every line has an ideal point, every line intersects the ideal line. It suffices to show that two ordinary lines intersect. The only possible problem could come from two lines whose real parts in $\mathbf{R}^2$ are parallel, but those will intersect in their common ideal point.

Theorem 3.4.2 is nice because arguments therefore do not have to contain special cases to deal with nonintersecting lines.

### 3.4.1 Analytic Properties of the Projective Plane

This section describes some analytic properties of $\mathbf{P}^2$.

**3.4.1.1. Theorem.** Three distinct points $[X_1,Y_1,Z_1]$, $[X_2,Y_2,Z_2]$, and $[X_3,Y_3,Z_3]$ of $\mathbf{P}^2$ are collinear if and only if

$$\begin{vmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \end{vmatrix} = 0. \tag{3.17}$$

**Proof.** We basically have to find numbers a, b, and c, not all zero, so that

$$aX_i + bY_i + cZ_i = 0, \quad i = 1,2,3.$$

The theorem is now an easy consequence of basic facts about when such systems of equations admit nontrivial solutions.

**3.4.1.2. Corollary.** The line in $\mathbf{P}^2$ determined by two distinct points $[X_1,Y_1,Z_1]$ and $[X_2,Y_2,Z_2]$ has equation

$$\begin{vmatrix} Y_1 & Z_1 \\ Y_2 & Z_2 \end{vmatrix} X + \begin{vmatrix} Z_1 & X_1 \\ Z_2 & X_2 \end{vmatrix} Y + \begin{vmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{vmatrix} Z = 0. \tag{3.18}$$

**Proof.** Simply apply Theorem 3.4.1.1 to points $[X,Y,Z]$, $[X_1,Y_1,Z_1]$, and $[X_2,Y_2,Z_2]$ and expand the determinant in the theorem by minors using the top row of the matrix.

**3.4.1.3. Theorem.** If the lines $\mathbf{L}_i$ in $\mathbf{P}^2$ defined by equations

$$a_iX + b_iY + c_iZ = 0, \quad i = 1,2,$$

are distinct, then they intersect in the point $[(a_1,b_1,c_1) \times (a_2,b_2,c_2)]$.

**Proof.** This follows from the fact that the cross product of two vectors is orthogonal to both of the vectors.

**3.4.1.4. Theorem.** Let $\mathbf{L}$ be the line in $\mathbf{P}^2$ determined by two distinct points $\mathbf{P}_1 = [\mathbf{p}_1]$ and $\mathbf{P}_2 = [\mathbf{p}_2]$, $\mathbf{p}_i \in \mathbf{R}^3$.

(1) Every point of $\mathbf{L}$ can be written in the form $\mathbf{P} = [s\mathbf{p}_1 + t\mathbf{p}_2]$ for some real numbers s and t that are not both zero. Furthermore, $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}$ are distinct points if $st \neq 0$. If $\mathbf{P} = [s'\mathbf{p}_1 + t'\mathbf{p}_2]$, then $(s',t') = c(s,t)$, for some nonzero constant c.

(2) Every point of **L** other than $\mathbf{P}_1$ can be expressed **uniquely** in the form $[s\mathbf{p}_1 + \mathbf{p}_2]$. By allowing s to equal ∞ and using the convention that $[s\mathbf{p}_1 + \mathbf{p}_2] = \mathbf{P}_1$ when $s = \infty$, we shall say in the future that **every** point of **L** can be so expressed. Similarly, every point of **L** can be expressed uniquely in the form $[\mathbf{p}_1 + t\mathbf{p}_2]$, where the point for $t = \infty$ is identified with $\mathbf{P}_2$.

**Proof.** The fact that every point $\mathbf{Q} = [\mathbf{q}]$ of **L** can be written in the form $[s\mathbf{p}_1 + t\mathbf{p}_2]$ follows form the fact that the vector **q** must be a linear combination of the vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ because the determinant (3.17) is zero. Showing that $\mathbf{P}_1$, $\mathbf{P}_2$, and **P** are distinct points if $st \neq 0$ is straightforward. Applying Theorem 3.4.1.1 to those three points implies that they lie on **L** because the fact that the bottom row is a linear combination of the top two rows means that the determinant in the theorem is zero. This basically proves (1). The existence part of (2) follows easily from (1). The uniqueness part of (2) is easily checked.

Note that identifying $[s\mathbf{p}_1 + \mathbf{p}_2]$ with $\mathbf{P}_1$ when $s = \infty$ is reasonable because

$$[s\mathbf{p}_1 + \mathbf{p}_2] = \left[s\left(\mathbf{p}_1 + \frac{1}{s}\mathbf{p}_2\right)\right] = \left[\mathbf{p}_1 + \frac{1}{s}\mathbf{p}_2\right] \to [\mathbf{p}_1] \quad \text{as} \quad s \to \infty.$$

The same argument justifies identifying $[\mathbf{p}_1 + t\mathbf{p}_2]$ with $\mathbf{P}_2$ when $t = \infty$.

Theorem 3.4.1.4(2) can be interpreted as saying that every line in $\mathbf{P}^2$ can be parameterized by the extended reals $\mathbf{R}^*$. In the first parameterization we have in effect made $\mathbf{P}_1$ the new ideal point. $\mathbf{P}_2$ is the new ideal point for the second. The parameterizations are not unique because they depend on the choice the two distinct points with respect to which they are defined. More importantly, the next example shows that it also depends on the **representatives** $\mathbf{p}_1$ and $\mathbf{p}_2$ we have chosen.

**3.4.1.5. Example.** Consider the line **L** defined by the points $\mathbf{P}_1 = [1,3,2]$ and $\mathbf{P}_2 = [-1,0,4]$. By Theorem 3.4.1.4(2),

$$\mathbf{L} = \{[s-1, 3s, 2s+4] \mid s \in \mathbf{R}\} \cup \{[-1,0,4]\}.$$

The point $\mathbf{Q} = [0,3,6]$ on **L** would be assigned the parameter $s = 1$. On the other hand, $\mathbf{P}_1 = [2,6,4]$ and choosing the representative (2,6,4) for $\mathbf{P}_1$ would have represented **L** as

$$\{[2s-1, 6s, 4s+4] \mid s \in \mathbf{R}\} \cup \{[-1,0,4]\},$$

and assigned the parameter 1/2 to **Q**.

Next, we would like to define the cross-ratio for four points on a line in $\mathbf{P}^2$. One approach would be to take advantage of our earlier definition of the cross-ratio for points in $\mathbf{R}^2$. The only complication is that one of the points might be an ideal point and one would have to give special definitions in those cases. It would be nice to give a more intrinsic definition. Although a metric-free definition was given by C.G. von Staudt (*Beiträge zur Geometrie der Lage*, 1847), the modern approach is based on metric considerations. We shall use a coordinate-based approach using homogeneous coordinates. In the end, one would of course want to check that all definitions agree.

An inherent problem when using coordinates is that one has to show that different coordinatizations give the same answer.

Let **A**, **B**, **C**, and **D** be four distinct points on a line **L** in $\mathbf{P}^2$. We shall denote the *cross-ratio in which **B** and **C** divide **A** and **D*** by (**AD**,**BC**). Here are two definitions for it:

**Definition 1.** Assume that **A** = [**a**] and **D** = [**d**] and express **B** and **C** in the form [**a** + k**d**] and [**a** + k′**d**], respectively, for some nonzero k and k′. Define

$$(\mathbf{AD},\mathbf{BC}) = \frac{k}{k'}. \tag{3.19}$$

**Definition 2.** Let **P** = [**p**] and **Q** = [**q**] be any distinct points of **L** and express **A**, **B**, **C**, and **D** in the form [**p** + a**q**], [**p** + b**q**], [**p** + c**q**], and [**p** + d**q**], respectively. Define

$$(\mathbf{AD},\mathbf{BC}) = \frac{(a-b)(d-c)}{(a-c)(d-b)}. \tag{3.20}$$

**3.4.1.6. Theorem.** The two definitions in equations (3.19) and (3.20) of the cross-ratio (**AD**,**BC**) of four distinct points **A**, **B**, **C**, and **D** on a line in the projective plane are well defined and agree.

**Proof.** The proof consists of some straightforward computations.

To prove that Definition 1 is well defined, let **A** = [s**a**] and **D** = [t**d**] and let **B** = [(s**a**) + m(t**d**)] and **C** = [(s**a**) + m′(t**d**)]. We must show that

$$\frac{m}{m'} = \frac{k}{k'}.$$

Now, (s**a**) + m(t**d**) = e(**a** + k**d**) and (s**a**) + m′(t**d**) = f(**a** + k′**d**) for nonzero constants e and f. It follows that

$$(s-e)\mathbf{a} = (ek-mt)\mathbf{d} \quad \text{and} \quad (s-f)\mathbf{a} = (fk'-m't)\mathbf{d}.$$

Since the vectors **a** and **d** are linearly independent, s = e = f, ek = mt, and fk′ = m′t, which easily implies that our two ratios are the same.

To prove that Definition 2 is well defined, one needs to show that it depends neither on the choice of **P** and **Q** nor on their representatives. The independence of their representatives is easy to show. For the rest, first replace **P** by **P′** = [**p′** ], **p′** = **p** + s**q**, express the points **A**, **B**, **C**, and **D** in terms of **P′** and **Q** as [**p′** + a′**q**], [**p′** + b′**q**], [**p′** + c′**q**], and [**p′** + d′**q**], respectively, and compare the new expressions with the old ones. One will find that a = s + a′, b = s + b′, etc., so that it is easy to check that

$$\frac{(a'-b')(d'-c')}{(a'-c')(d'-b')} = \frac{(a-b)(d-c)}{(a-c)(d-b)}.$$

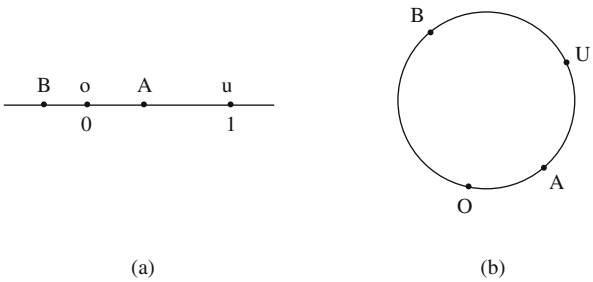



**Figure 3.10.** Defining when a point is between two points.

Similarly, one shows that replacing **Q** by **Q′** will preserve formula (3.20). Let **Q′** = [**q′**], **q′** = **p** + t**q**. Express the points **A**, **B**, **C**, and **D** in terms of **P** and **Q′** and relate the new expressions with the old ones. The details are left to the reader.

Finally, to show that Definition 1 and 2 agree, choose **P** = **A** and **Q** = **D**. In this case, a = 0 and d = ∞. These values imply that the quotient in (3.20) is just b/c, the same value that Definition 1 asserts. This proves the theorem.

We shall give a third definition of the cross-ratio shortly. First, we study the problem of coordinatizing the points of the projective line and plane. Let us start with lines. The standard way that one assigns coordinates (real numbers) to a line in Euclidean space (thought of as an abstract set of vectors without coordinates) is to decide on a unit of distance, pick a start point **o** (the "zero"), and pick a direction for the line that defines which half with respect to **o** will get positive numbers and which will get negative numbers. We can accomplish this by picking two points: the start point **o** and another point **u** (the "+1") that is a unit distance form **o**. See Figure 3.10(a). Each point **p** of the line is then assigned the number t, where **p** − **o** = t(**u** − **o**). Note how **o** and **u** get assigned the numbers 0 and 1, respectively. What is different about a line in $\mathbf{P}^2$ is that it is topologically a circle. Picking two points on a circle does not orient it. In Figure 3.10(b), which of the points **A** or **B** is "between" **O** and **U**? Figure 3.10(a) shows that this is not a problem for a line in $\mathbf{R}^2$. For a circle or projective line we have to pick three points, but it is convenient to pick them in a special way. This will also restore the dependency on representations of points that we lost in Theorem 3.4.1.4. The next lemma is the basic property we need.

**3.4.1.7. Lemma.** Let **I** = [$\mathbf{i}_1$] = [$\mathbf{i}_2$], **O** = [$\mathbf{o}_1$] = [$\mathbf{o}_2$], and **U** = [$\mathbf{u}_1$] = [$\mathbf{u}_2$] be three distinct points on a line in $\mathbf{P}^2$. If

$$\mathbf{i}_j + \mathbf{o}_j = \mathbf{u}_j, \quad j = 1,2, \tag{3.21}$$

then $\mathbf{i}_1 = c\mathbf{i}_2$, $\mathbf{o}_1 = c\mathbf{o}_2$, and $\mathbf{u}_1 = c\mathbf{u}_2$, for some $c \neq 0$.

**Proof.** Let $\mathbf{i}_1 = a\mathbf{i}_2$, $\mathbf{o}_1 = b\mathbf{o}_2$, and $\mathbf{u}_1 = c\mathbf{u}_2$. Equations (3.21) imply that

$$a\mathbf{i}_2 + b\mathbf{o}_2 = c\mathbf{u}_2 = c(\mathbf{i}_2 + \mathbf{o}_2) = c\mathbf{i}_2 + c\mathbf{o}_2.$$

Since the vectors $\mathbf{i}_2$ and $\mathbf{o}_2$ are linearly independent, we must have a = b = c.

The importance of Lemma 3.4.1.7 is the following:

**3.4.1.8. Corollary.** Using the notation in Theorem 3.4.1.4, if we require the representatives $\mathbf{p}_1$ and $\mathbf{p}_2$ to always add up to a representative of a fixed point, then the parameterization described in part (2) of that theorem will depend **only** on the points $\mathbf{P}_1$ and $\mathbf{P}_2$ but not on their representatives.

Consider the projective line $\mathbf{P}^1$ and the three special points [1,0], [0,1], and [1,1] that correspond to the points ∞, 0, and 1. Note how the coordinates of the standard representatives of the first two points add up to the coordinates of the representative of the third. We are ready to coordinatize points in projective space.

**Definition.** Let $\mathbf{L}$ be a line in $\mathbf{P}^2$ and let $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{U}$ be three distinct points on $\mathbf{L}$. Choose representations $\mathbf{I} = [\mathbf{i}]$, $\mathbf{O} = [\mathbf{o}]$, and $\mathbf{U} = [\mathbf{u}]$ for the points so that $\mathbf{i} + \mathbf{o} = \mathbf{u}$. The map

$$\varphi : \mathbf{P}^1 \to \mathbf{L}$$

defined by

$$\varphi([X, Y]) = [X\mathbf{i} + Y\mathbf{o}]$$

is called the *standard parameterization of* $\boldsymbol{L}$ *with respect to the points* $\boldsymbol{I}$, $\boldsymbol{O}$, *and* $\boldsymbol{U}$. Using the standard identification of $\mathbf{P}^1$ and $\mathbf{R}^*$ we shall also describe the map φ with the formulas

$$\varphi(x) = [x\mathbf{i} + \mathbf{o}] \quad \text{and} \quad \varphi(\infty) = \mathbf{I},$$

**3.4.1.9. Theorem.** The standard parameterization of a line $\mathbf{L}$ with respect to three of its points is a one-to-one and onto map that depends only on the points and not on their representatives.

**Proof.** The theorem follows from Corollary 3.4.1.8.

**Definition.** Let φ be the standard parameterization of a line $\mathbf{L}$ in $\mathbf{P}^2$ with respect to points $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{U}$. If $\mathbf{P} \in \mathbf{L}$ and if $\varphi^{-1}(\mathbf{P}) = [X,Y]$, then (X,Y) will be called the *homogeneous coordinates* of $\mathbf{P}$ with respect to the *coordinate system* defined by $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{U}$. Let $x = X/Y$ or ∞ depending on whether $Y \neq 0$ or $Y = 0$. The number x will be called the (*extended real* or *affine*) *coordinate* of $\mathbf{P}$ with respect to the given coordinate system. The points [1,0], [0,1], and [1,1] define the *standard coordinate system* for $\mathbf{P}^1$ and the coordinates with respect to it are called the *standard coordinates*.

Let

$$\varphi_k : \mathbf{P}^1 \to \mathbf{L} \tag{3.22a}$$

be the standard parameterization of a line $\mathbf{L}$ with respect to the coordinate system defined by points $\mathbf{I}_k$, $\mathbf{O}_k$, and $\mathbf{U}_k$ on $\mathbf{L}$, k = 1,2. Express $\mathbf{I}_k$, $\mathbf{O}_k$, and $\mathbf{U}_k$ in the form $\mathbf{I}_k = [\mathbf{i}_k]$, $\mathbf{O}_k = [\mathbf{o}_k]$, and $\mathbf{U}_k = [\mathbf{u}_k]$ with $\mathbf{i}_k + \mathbf{o}_k = \mathbf{u}_k$. By Theorem 3.4.1.4 there are constants a, b, c, and d, so that

$$\mathbf{i}_1 = a\mathbf{i}_2 + c\mathbf{o}_2,$$
$$\mathbf{o}_1 = b\mathbf{i}_2 + d\mathbf{o}_2. \tag{3.22b}$$

**3.4.1.10. Theorem.** Given the parameterizations $\varphi_k$ in (3.22a) and the constants a, b, c, and d in equations (3.22b), the map

$$\psi = \varphi_2^{-1}\varphi_1\colon\ \mathbf{P}^1 \to \mathbf{P}^1$$

has the form

$$\psi([X,Y]) = [aX + bY, cX + dY] \quad \text{with} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0. \tag{3.22c}$$

If we identify $\mathbf{P}^1$ with $\mathbf{R}^*$ and consider $\Psi$ as a map from $\mathbf{R}^*$ to $\mathbf{R}^*$, then the map $\Psi$ has on the form

$$\psi(x) = \frac{ax + b}{cx + d}. \tag{3.22d}$$

**Proof.** The theorem follows from the following string of equalities:

$$\begin{aligned}\varphi_1([X,Y]) &= [X\mathbf{i}_1 + Y\mathbf{o}_1] \\ &= [X(a\mathbf{i}_2 + c\mathbf{o}_2) + Y(b\mathbf{i}_2 + d\mathbf{o}_2)] \\ &= [(aX + bY)\mathbf{i}_2 + (cX + dY)\mathbf{o}_2] \\ &= \varphi_2([aX + bY, cX + dY]).\end{aligned}$$

The determinant in (3.22c) is nonzero because $\mathbf{i}_1$ and $\mathbf{o}_1$ are linearly independent. Equation (3.22d) is obtained by factoring out a Y and making the substitution x = X/Y.

Theorem 3.4.1.10 should be taken as a statement about how coordinates change as one moves from one coordinate system for a projective line to another. Specifically, we have

**3.4.1.11. Corollary.** Using the notation in Theorem 3.4.1.10, the map

$$(X\ Y) \to (X\ Y)\begin{pmatrix} a & c \\ b & d \end{pmatrix} \quad \text{with} \quad \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0, \tag{3.23}$$

maps the homogeneous coordinates (X,Y) of a point of the line **L** with respect to the coordinate system defined by $\mathbf{I}_1$, $\mathbf{O}_1$, and $\mathbf{U}_1$ to the homogeneous coordinates of the same point with respect to the coordinate system defined by $\mathbf{I}_2$, $\mathbf{O}_2$, and $\mathbf{U}_2$. Conversely, every such map corresponds to a change of coordinates.

**3.4.1.12. Example.** Suppose the standard coordinates for a point **P** in $\mathbf{P}^1$ are 5 = [5,1]. What are the coordinates of **P** with respect to the coordinate system defined by **I** = 2 = [2,1], **O** = 3 = [3,1], and **U** = 7 = [7,1]?

**Solution.** First of all, we must find representatives **i** and **o** for **I** and **O**, respectively, so that **i** + **o** = (7,1). Since the equation

$$a(2,1) + b(3,1) = (7,1)$$

has solution a = –4 and b = 5, we can let **i** = (–8,–4) and **o** = (15,5). Next, the standard parameterization φ of $\mathbf{P}^1$ with respect to the given **I**, **O**, and **U** is defined by

$$\varphi(x) = [x(-8,-4) + (15,5)] = [-8x + 15, -4x + 5] = \frac{-8x+15}{-4x+5}.$$

Since φ(5/6) = 5, it follows that 5/6 are the coordinates of **P** in the new coordinate system. Alternatively, we could follow the proof of Theorem 3.4.1.10 and solve the system of equations

$$(1,0) = a(-8,-4) + c(15,5)$$
$$(0,1) = b(-8,-4) + d(15,5)$$

for a, b, c, and d to get a = 1/4, b = –3/4, c = 1/5, and d = –2/5. Then

$$(5\ \ 1)\begin{pmatrix} \frac{1}{4} & \frac{1}{5} \\ -\frac{3}{4} & -\frac{2}{5} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{3}{5} \end{pmatrix} \leftrightarrow \frac{5}{6}$$

which leads to the same answer.

We are ready for the third definition of the cross-ratio (**AD**,**BC**) of four distinct points **A**, **B**, **C**, and **D** on a line **L** in $\mathbf{P}^2$.

**Definition 3.** Let φ be the standard parameterization of **L** with respect to a coordinate system **I**, **O**, and **U**. If $\varphi([a_1,a_2]) = \mathbf{A}$, $\varphi([b_1,b_2]) = \mathbf{B}$, $\varphi([c_1,c_2]) = \mathbf{C}$, and $\varphi([d_1,d_2]) = \mathbf{D}$, then define

$$(\mathbf{AD},\mathbf{BC}) = \frac{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\begin{vmatrix} d_1 & d_2 \\ c_1 & c_2 \end{vmatrix}}{\begin{vmatrix} d_1 & d_2 \\ b_1 & b_2 \end{vmatrix}\begin{vmatrix} a_1 & a_2 \\ c_1 & c_2 \end{vmatrix}}. \tag{3.24}$$

**3.4.1.13. Theorem.** The definition of the cross-ratio (**AD**,**BC**) of four distinct points **A**, **B**, **C**, and **D** on a line **L** in the projective plane via equation (3.24) is well defined and agrees with those in equations (3.19) and (3.20).

**Proof.** To show that the definition is independent of the coordinate system one only needs to show that the right-hand side of equation (3.24) is unchanged under a transformation of the form in equation (3.23). A straightforward computation does that. Next, it suffices to show agreement with Definition 1. Let φ be the standard parame-

terization of **L** with respect to the points **A** = [**a**], **D** = [**d**], and **C** = [**c**] with **a** + **d** = **c**. Then $\varphi(1,0) = \mathbf{A}$, $\varphi(0,1) = \mathbf{D}$, $\varphi(1,1) = \mathbf{C}$, and the right-hand side of equation (3.24) evaluates to $b_2/b_1$. Now, $\varphi(b_1,b_2) = \mathbf{B} = [\mathbf{a} + k\mathbf{d}] = \varphi(1,k)$, which implies that $k = b_2/b_1$. In the notation of Definition 1, the cross-ratio (**AD**,**BC**) was defined to be $k/k'$, but $k' = 1$ here, so that the theorem is proved.

For yet another derivation and explanation of the projective invariance of the cross-ratio see [Blin98].

We now move on to parameterizing points in the projective plane. What we did for a projective line will generalize in a natural way. Three points were enough to determine a coordinate system for $\mathbf{R}^2$. We need a fourth for $\mathbf{P}^2$.

**Definition.** Let **I**, **J**, **O**, and **U** be four points of $\mathbf{P}^2$ no three of which are collinear. Choose representations **I** = [**i**], **J** = [**j**], **O** = [**o**], and **U** = [**u**] for the points so that **i** + **j** + **o** = **u**. The map

$$\varphi : \mathbf{P}^2 \to \mathbf{P}^2$$

defined by

$$\varphi([X,Y,Z]) = [X\mathbf{i} + Y\mathbf{j} + Z\mathbf{o}]$$

is called the *standard parameterization of* $\boldsymbol{P}^2$ *with respect to the coordinate system* ***I***, ***J***, ***O***, *and* ***U***. Using the standard inclusion of $\mathbf{R}^2$ in $\mathbf{P}^2$ we shall also describe the map $\varphi$ with the formulas

$$\varphi(x,y) = [x\mathbf{i} + y\mathbf{j} + \mathbf{o}] \quad \text{and} \quad \varphi([x,y,0]) = [x\mathbf{i} + y\mathbf{j}],$$

**3.4.1.14. Theorem.** The standard parameterization of $\mathbf{P}^2$ with respect to four of its points is a one-to-one and onto map that depends only on the points and not on their representatives.

**Proof.** The theorem follows from the natural analogs of Theorem 3.4.1.4, Lemma 3.4.1.7, and Corollary 3.4.1.8, whose proofs are left as exercises for the reader.

**Definition.** Let $\varphi$ be the standard parameterization of $\mathbf{P}^2$ with respect to points **I**, **J**, **O**, and **U**. If $\mathbf{P} \in \mathbf{P}^2$ and if $\varphi^{-1}(\mathbf{P}) = [X,Y,Z]$, then (X,Y,Z) will be called the *homogeneous coordinates* of **P** with respect to the *coordinate system* defined by **I**, **J**, **O**, and **U**. For those points **P** for which $Z \neq 0$, let $x = X/Z$ and $y = Y/Z$ and call the pair (x,y) the (*affine*) *coordinates* of **P** with respect to the given coordinate system. The points [1,0,0], [0,1,0], [0,0,1], and [1,1,1] define the *standard coordinate system* for $\mathbf{P}^2$ and the coordinates with respect to it are called the *standard coordinates*.

**3.4.1.15. Example.** Suppose the standard coordinates for a point **P** in $\mathbf{P}^2$ are (1,2) = [1,2,1]. What are the coordinates of **P** with respect to the coordinate system defined by **I** = (–3,0) = [–3,0,1], **J** = (0,–2) = [0,–2,1], **O** = (–1,–1) = [–1,–1,1], and **U** = (–2,–1) = [–2,–1,1]?

**Solution.** The representatives $\mathbf{i} = (-3,0,1)$, $\mathbf{j} = (0,-2,1)$, and $\mathbf{o} = (1,1,-1)$ for $\mathbf{I}$, $\mathbf{J}$, and $\mathbf{O}$, respectively, have the property that they add up to a representative for $\mathbf{U}$. The standard parameterization $\varphi$ of $\mathbf{P}^2$ with respect to the given $\mathbf{I}$, $\mathbf{J}$, $\mathbf{O}$, and $\mathbf{U}$ is defined by

$$\begin{aligned}\varphi(x,y) &= [x(-3,0,1) + y(0,-2,1) + (1,1,-1)] \\ &= [(-3x+1,-2y+1,1)] \qquad (\leftrightarrow (-3x+1,-2y+1)).\end{aligned}$$

Solving the equation $\varphi(x,y) = (1,2)$ for x and y gives that $x = 0$ and $y = -1/2$. Therefore, $\mathbf{P}$ has coordinates $(0,-1/2)$ with respect to the given coordinate system.

Let

$$\varphi_k : \mathbf{P}^2 \to \mathbf{P}^2, \tag{3.25a}$$

be the standard parameterization of $\mathbf{P}^2$ with respect to the coordinate system defined by $\mathbf{I}_k$, $\mathbf{J}_k$ $\mathbf{O}_k$, and $\mathbf{U}_k$, $k = 1,2$. Express $\mathbf{I}_k$, $\mathbf{J}_k$ $\mathbf{O}_k$, and $\mathbf{U}_k$ in the form $\mathbf{I}_k = [\mathbf{i}_k]$, $\mathbf{J}_k = [\mathbf{j}_k]$, $\mathbf{O}_k = [\mathbf{o}_k]$, and $\mathbf{U}_k = [\mathbf{u}_k]$ with $\mathbf{i}_k + \mathbf{j}_k + \mathbf{o}_k = \mathbf{u}_k$. Because the vectors $\mathbf{i}_k$, $\mathbf{j}_k$, and $\mathbf{o}_k$ are linearly independent, there are constants $a_i$, $b_i$, and $c_i$, so that

$$\begin{aligned}\mathbf{i}_1 &= a_1\mathbf{i}_2 + b_1\mathbf{j}_2 + c_1\mathbf{o}_2, \\ \mathbf{j}_1 &= a_2\mathbf{i}_2 + b_2\mathbf{j}_2 + c_2\mathbf{o}_2, \\ \mathbf{o}_1 &= a_3\mathbf{i}_2 + b_3\mathbf{j}_2 + c_3\mathbf{o}_2.\end{aligned} \tag{3.25b}$$

**3.4.1.16. Theorem.** Given the parameterizations $\varphi_k$ in (3.25a) and the constants $a_i$, $b_i$, and $c_i$ in equations (3.25b), the map

$$\psi = \varphi_2^{-1}\varphi_1 : \mathbf{P}^2 \to \mathbf{P}^2$$

has the form

$$\psi([X,Y,Z]) = [a_1X + a_2Y + a_3Z, b_1X + b_2Y + b_3Z, c_1X + c_2Y + c_3Z], \tag{3.25c}$$

$$\text{with} \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0.$$

With respect to the standard inclusion of $\mathbf{R}^2$ in $\mathbf{P}^2$ the map $\Psi$ has on the form

$$\psi(x,y) = \left(\frac{a_1x + a_2y + a_3}{c_1x + c_2y + c_3}, \frac{b_1x + b_2y + b_3}{c_1x + c_2y + c_3}\right). \tag{3.25d}$$

**Proof.** The proof is a straightforward computation similar to the proof of Theorem 3.4.1.10.

Like Theorem 3.4.1.10, Theorem 3.4.1.16 should be taken as a statement about how coordinates change as one moves from one coordinate system of the projective plane to another. In other words, we have

**3.4.1.17. Corollary.** Using the notation in Theorem 3.4.1.16, the map

$$(X\ Y\ Z) \to (X\ Y\ Z)\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \quad \text{with} \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0 \tag{3.26}$$

maps the homogeneous coordinates (X,Y,Z) of a point of $\mathbf{P}^2$ with respect to the coordinate system defined by $\mathbf{I}_1$, $\mathbf{J}_1$, $\mathbf{O}_1$, and $\mathbf{U}_1$ to the homogeneous coordinates of the same point with respect to the coordinate system defined by $\mathbf{I}_2$, $\mathbf{J}_2$, $\mathbf{O}_2$, and $\mathbf{U}_2$. Conversely, every such map corresponds to a change of coordinates.

## 3.4.2 Two-dimensional Projective Transformations

This section defines the natural transformations associated to the projective plane and discusses some of their analytic properties. The next definition is in the spirit of the definitions of the affine transformations in Chapter 2 and the approach in Section 3.2.

**Definition.** A *projective transformation* or *projectivity* of $\mathbf{P}^2$ is any one-to-one and onto map $T : \mathbf{P}^2 \to \mathbf{P}^2$ that preserves collinearity and the cross-ratio of points.

Compare this definition with the one in Section 3.2 and note that we no longer have to add provisions about things being defined. Our definition has become much cleaner. However, to make working with such transformations easy we need to derive their analytic form.

Recall the fractional transformations of the plane defined in Section 3.2. Translating them into equations dealing with homogeneous rather than Cartesian coordinates leads to the following homogeneous system of equations:

$$\begin{aligned} x' &= a_1x + a_2y + a_3z \\ y' &= b_1x + b_2y + b_3z, \quad \text{where} \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0. \\ z' &= c_1x + c_2y + c_3z \end{aligned} \tag{3.27}$$

**3.4.2.1. Theorem.** The system of equations in (3.27) determines a well-defined one-to-one and onto transformation T: $\mathbf{P}^2 \to \mathbf{P}^2$, which preserves collinearity and the cross-ratio. In other words, the system in (3.27) determines a projectivity. Conversely, every projectivity of $\mathbf{P}^2$ can be described via a system of equations as in (3.27).

**Proof.** The first part of the theorem is fairly straightforward. For the second, see [Gans69].

**Definition.** The matrix

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

will be called *a matrix for the projective transformation* defined by (3.27). (We choose this matrix rather than the transpose because we shall let vectors operate on the left to be consistent with what we do with matrices for linear transformations.)

Note that the coefficients in the system of equations in (3.27) are **not** uniquely determined by the projective transformation. If one were to multiply every coefficient by some fixed nonzero number, the new coefficients would define the same transformation. This is why we defined **a** matrix for a projective transformation rather than **the** matrix. The entries of the matrix **are** unique up to such a common multiple however.

**3.4.2.2. Theorem.** The projectivities of $\mathbf{P}^2$ form a group under composition.

**Proof.** This is straightforward.

Next, let us relate the projectivities defined here with the transformations defined in Chapter 2. First of all, it is clear that the projectivities defined in Section 3.2 are just the projectivities of $\mathbf{P}^2$ restricted to $\mathbf{R}^2$.

**Definition.** An *affine transformation* of $\mathbf{P}^2$ is any projectivity T of $\mathbf{P}^2$ with the property that $T(\mathbf{R}^2) \subseteq \mathbf{R}^2$.

One can show that the set of affine transformations forms a subgroup of the group of projective transformations. Also, if T is an affine transformation, then T must necessarily send ideal points to ideal points. In other words, if $T([x,y,z]) = [x',y',z']$, then $z = 0$ implies that $z' = 0$. Using the notation in (3.27) this means that $c_1x + c_2y = 0$ for all x and y. The only way that this can be true is if $c_1 = c_2 = 0$. This shows that the equations for an affine transformation have the form

$$\begin{aligned} x' &= a_1x + a_2y + a_3z \\ y' &= b_1x + b_2y + b_3z, \\ z' &= \qquad\qquad c_3z \end{aligned} \quad \text{where} \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 0 & 0 & c_3 \end{vmatrix} \neq 0. \tag{3.28a}$$

and that they have matrices of the form

$$\begin{pmatrix} a_1 & b_1 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & c_3 \end{pmatrix} \tag{3.28b}$$

Note that we could have normalized the $c_3$ in (3.28a) and (3.28b) to be 1.

The equations show that the affine transformations of $\mathbf{P}^2$ are just the extensions of the affine transformations of $\mathbf{R}^2$ as defined in Chapter 2. (Simply translate the equations found in Chapter 2 into equations using homogeneous coordinates.) In particular, the similarities and motions of $\mathbf{R}^2$ extend to transformations of $\mathbf{P}^2$. The equations for motions in terms of homogeneous coordinates have the form

$$\begin{aligned} x' &= \phantom{-}ax + by + cz \\ y' &= -bx + ay + dz \\ z' &= \phantom{-bx + ay +{}} z, \end{aligned} \tag{3.29}$$

where $a^2 + b^2 = 1$.

Comparing the results of Section 3.2 with the corresponding results for the projective plane, it is clear that the projective plane has simplified the mathematics. Aside from theoretical gains though, of what practical value is all of this?

### 3.4.3 Planar Maps and Homogeneous Coordinates

This section shows how homogeneous coordinates can simplify defining maps in the plane, in particular central projections and perspectivities. The reason is that **all** (projective) transformations can be expressed uniformly via matrices. Matrices are convenient from the point of view of computations and it is advantageous to be able to represent the composition of transformations by matrix products. This is not possible with Cartesian coordinates. The problem is that translations cannot be expressed as matrices when one uses Cartesian coordinates.

Consider the projective transformation

$$U_a : \mathbf{P}^2 \to \mathbf{P}^2$$

defined by the matrix

$$M(U_a) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix}.$$

This is a very special projectivity that is closely related to the central projection of the plane onto the x-axis from the point (0,–1/a) on the y-axis. The latter sends the point (x,y) in the plane to the point (x/(ay + 1),0) on the x-axis. To see this, consider Figure 3.11 and note that by similarity of triangles

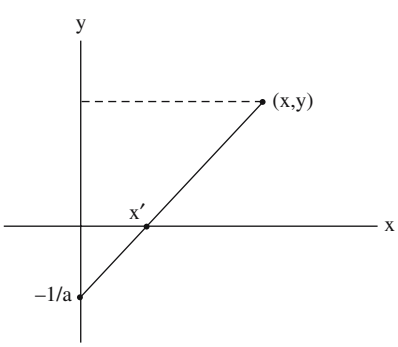



**Figure 3.11.** A standard central projection.

$$\frac{1/a}{y-(-1/a)} = \frac{x'}{x}, \quad \text{that is,} \quad x' = \frac{x}{ay+1}.$$

On the other hand,

$$U_a([x,y,1]) = [(x,y,1)\,M(U_a)] = [x,y,ay+1] = \left[\frac{x}{ay+1}, \frac{y}{ay+1}, 1\right].$$

This shows that the central projection

$$C_a : \mathbf{R}^2 \to \mathbf{R}$$

onto the x-axis from the point (0,–1/a) on the y-axis can be represented by the matrix

$$M_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & a \\ 0 & 0 & 1 \end{pmatrix},$$

as long as one uses the homogeneous coordinate representation for points. This means that one can also use matrix algebra to deal with central projections.

**3.4.3.1. Example.** Find the central projection C of the plane onto the line **L** defined by the equation x – y = 2 from the point **P** = (5,1). Show that C(5,4) = (5,3).

**Solution.** See Figure 3.12. Since we now know how compute central projection onto the x-axis from points on the y-axis, the idea will be to reduce this problem to one of that type. One way to achieve this situation is to translate **P** to (1,–1) and then to rotate about the origin through an angle of $-\pi/4$. Let T be this translation and R the rotation. Then $\mathbf{P}' = RT(\mathbf{P}) = (0,-\sqrt{2})$ and $C = T^{-1}R^{-1}SRT$, where $S = C_{1/\sqrt{2}}$ is the central

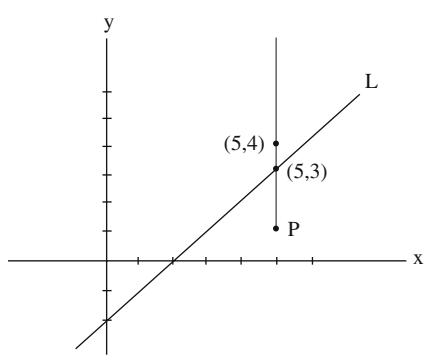



**Figure 3.12.** A central projection example.

projection from the point **P′** onto the x-axis. The matrices for these maps with respect to homogeneous coordinates are:

$$T \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -2 & 1 \end{pmatrix} \qquad T^{-1} \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 2 & 1 \end{pmatrix} \qquad S \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 1 \end{pmatrix}$$

$$R \leftrightarrow \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad R^{-1} \leftrightarrow \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Multiplying these matrices together gives that

$$C \leftrightarrow \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{5}{2} & \frac{3}{2} & \frac{1}{2} \\ 5 & 1 & 2 \end{pmatrix}.$$

In other words, C is defined by the equations

$$x' = \frac{-\frac{3}{2}x + \frac{5}{2}y + 5}{-\frac{1}{2}x + \frac{1}{2}y + 2}, \quad y' = \frac{-\frac{1}{2}x + \frac{3}{2}y + 1}{-\frac{1}{2}x + \frac{1}{2}y + 2}.$$

Since

$$(5 \;\; 4 \;\; 1)\begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{5}{2} & \frac{3}{2} & \frac{1}{2} \\ 5 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{15}{2} & \frac{9}{2} & \frac{3}{2} \end{pmatrix},$$

and

$$\left[\frac{15}{2}, \frac{9}{2}, \frac{3}{2}\right] = [5,3,1],$$

it follows that C(5,4) = (5,3). We could also have deduced this by substituting 5 for x and 3 for y in the equations above.

Frames could also be used to solve problems like the one in Example 3.4.3.1. See Section 3.5.1.

There is another interesting aspect to the projectivity $U_a$. Consider the lines

$$mx + y = -\frac{1}{a} \quad \text{and} \quad mx - y = \frac{1}{a}$$

through (0,–1/a) and with slope m. Note that

$$(x, y, z)\, M_a = (x, y, ay + z).$$

In particular,

$$\left(0, -\frac{1}{a}, 1\right) M_a = \left(0, -\frac{1}{a}, 0\right),$$

$$\left(x, -\frac{1}{a} - mx, 1\right) M_a = \left(x, -\frac{1}{a} - mx, -amx\right) = -\frac{1}{amx}\left(-\frac{1}{am}, \frac{1}{a} + \frac{1}{a^2mx}, 1\right), \quad \text{and}$$

$$\left(x, -\frac{1}{a} + mx, 1\right) M_a = \left(x, -\frac{1}{a} + mx, amx\right) = \frac{1}{amx}\left(\frac{1}{am}, \frac{1}{a} - \frac{1}{a^2mx}, 1\right).$$

This shows that the center (0,–1/a) of the central projection has been mapped to "infinity" and the two lines have been mapped to the lines

$$x' = -\frac{1}{am} \quad \text{and} \quad x' = \frac{1}{am},$$

respectively. See Figure 3.13. In general, lines through (0,–1/a) are mapped to vertical lines through their intersection with the x-axis. Furthermore, the central projection from (0,–1/a) has been transformed into an orthogonal projection of $\mathbf{R}^2$ onto the x-axis. Dually, the vertical lines x = c are mapped to lines through (c,0), which inter-

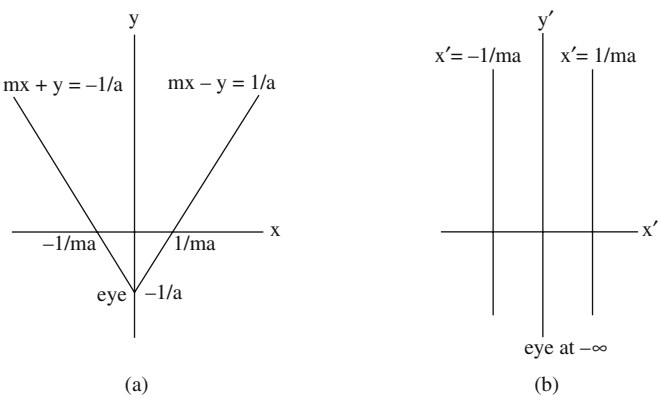



**Figure 3.13.** The geometry of a central projection.

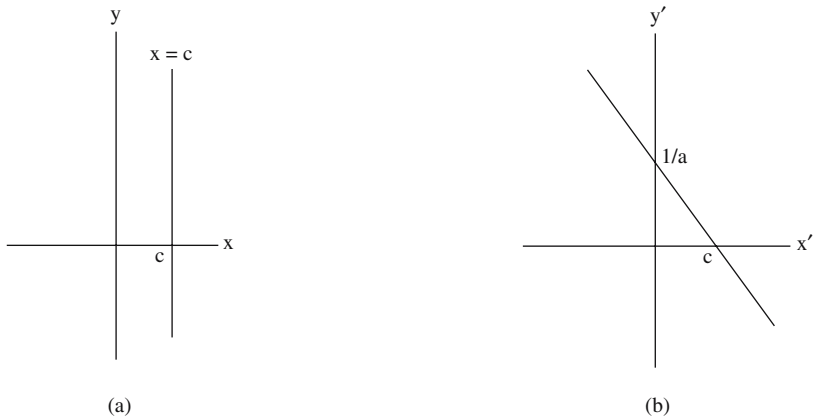



**Figure 3.14.** A vanishing point.

sect in (0,1/a). See Figure 3.14. The point (0,1/a) is called a *vanishing point*. The reader should recall the discussion of vanishing points in Section 3.2.

The last observation about the map $U_a$ is that it is a well-defined map on all of projective space even though the affine central projection with center (0,–1/a) is not, since it is not defined for the points on the line y = –1/a. The simple-minded explanation for this discrepancy is that affine central projections involve a division, so that one has to worry about zero denominators, whereas the associated projective transformations do not.

Finally, consider the projective transformation $U_{a,b} : \mathbf{P}^2 \to \mathbf{P}^2$ defined by the matrix

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.30}$$

The affine map T that it induces also has a simple geometric description. Let $T_1$ be the parallel projection in direction $\mathbf{e}_3$ of the x-y plane onto the plane $\mathbf{X}$ defined by

$$ax + by - z = 0.$$

Let $T_2$ be the perspectivity with center (0,0,–1) from $\mathbf{X}$ back to the x-y plane. It is easy to show that $T = T_2T_1$. See Figure 3.15. The point $\mathbf{P}$ in the figure gets mapped to $\mathbf{Q}$ in the plane $\mathbf{X}$ by $T_1$ and then to $\mathbf{P}'$ by $T_2$.

## 3.5 Beyond the Plane

Central projections, perspectivities, projective transformations, fractional transformations, and homogeneous coordinates can all be defined in a straight forward

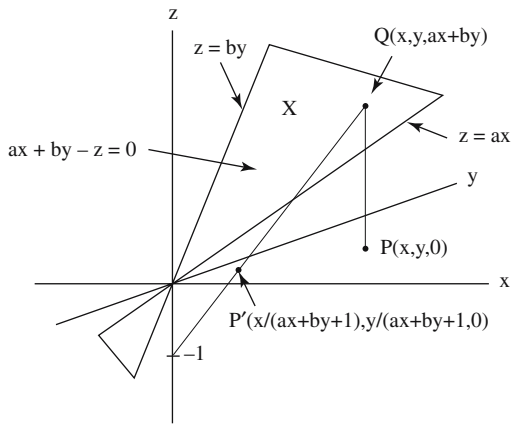



**Figure 3.15.** A geometric explanation of the transformation defined by (3.30).

manner for $\mathbf{R}^n$, but restricting ourselves to $\mathbf{R}^n$ would lead to the same shortcomings that we saw for $\mathbf{R}^2$. The proper setting for all these maps is again projective space. N-dimensional projective space $\mathbf{P}^n$ was defined in Section 3.4. This section briefly describes how the definitions and results of the last few sections extend.

First of all, in analogy with $\mathbf{P}^2$, one can think of $\mathbf{P}^n$ as the set of lines through the origin in $\mathbf{R}^{n+1}$. Furthermore, there is again a natural imbedding of $\mathbf{R}^n$ into $\mathbf{P}^n$.

**Definition.** The map

$$\iota : \mathbf{R}^n \to \mathbf{P}^n$$

defined by

$$\iota(x_1, x_2, \ldots, x_n) = [x_1, x_2, \ldots, x_n, 1]$$

is called the *standard imbedding of* $\boldsymbol{R}^n$ *in* $\boldsymbol{P}^n$.

Identifying a point $\mathbf{p}$ of Euclidean space with its image $\iota(\mathbf{p})$ in projective space allows us to consider $\mathbf{R}^n$ as a subset of $\mathbf{P}^n$. We shall do so from now on and consider $\iota$ as an inclusion map. Also, by identifying $[x_1, x_2, \ldots, x_i, x_{i+1}]$ in $\mathbf{P}^i$ with $[x_1, x_2, \ldots, x_i, 0, x_{i+1}]$ in $\mathbf{P}^{i+1}$, we get *natural inclusions* $\mathbf{P}^i \subset \mathbf{P}^{i+1}$ and a commutative diagram

$$\begin{array}{ccccccccc}
\mathbf{P}^0 & \subset & \mathbf{P}^1 & \subset & \mathbf{P}^2 & \subset \ldots \subset & \mathbf{P}^{n-1} & \subset & \mathbf{P}^n \\
\| & & \cup & & \cup & & \cup & & \cup \\
\mathbf{R}^0 & \subset & \mathbf{R}^1 & \subset & \mathbf{R}^2 & \subset \ldots \subset & \mathbf{R}^{n-1} & \subset & \mathbf{R}^n.
\end{array}$$

**Definition.** The points of $\mathbf{P}^n - \mathbf{R}^n$ are called *ideal points*. All the other points are called *real points*.

The ideal points in $\mathbf{P}^n$ are the points of the form $[x_1, \ldots, x_n, 0]$. They can also be thought of as corresponding to families of parallel lines in $\mathbf{R}^n$. Lines, planes, etc., in $\mathbf{P}^n$ are defined in terms of solutions to appropriate sets of homogeneous equations.

**Definition.** A (*projective*) *hyperplane* in $\mathbf{P}^n$ is any subset of $\mathbf{P}^n$ of the form

$$\{[X_1, X_2, \cdots, X_{n+1}] \mid a_1X_1 + a_2X_2 + \cdots + a_{n+1}X_{n+1} = 0, \text{ for fixed } (a_1, a_2, \ldots, a_{n+1}) \neq \mathbf{0}\}.$$

**Definition.** A *k-dimensional projective plane* in $\mathbf{P}^n$ is any subset of $\mathbf{P}^n$ of the form

$$\begin{aligned}\{[X_1, X_2, \ldots, X_{n+1}] \mid \quad & a_{11}X_1 + \quad a_{12}X_2 + \cdots + \quad a_{1,n+1}X_{n+1} = 0,\\ & a_{21}X_1 + \quad a_{22}X_2 + \cdots + \quad a_{2,n+1}X_{n+1} = 0,\\ & \quad . \qquad\qquad . \qquad\qquad\qquad . \\ & a_{n-k,1}X_1 + a_{n-k,2}X_2 + \cdots + a_{n-k,n+1}X_{n+1} = 0\}\end{aligned}$$

where $(a_{i1}, a_{i2}, \ldots, a_{i,n+1})$, $i = 1, \ldots, n - k$, are a fixed set of $n - k$ linearly independent vectors. A one-dimensional projective plane in $\mathbf{P}^n$ is called a (*projective*) *line*.

One can show that a projective hyperplane in $\mathbf{P}^n$ is just an $(n - 1)$-dimensional projective plane. A k-dimensional projective plane should be thought of as an imbedded $\mathbf{P}^k$ (see Corollary 3.5.2 below). The ideal points in $\mathbf{P}^n$ form a hyperplane defined by the equation

$$X_{n+1} = 0.$$

A projective line either consists entirely of ideal points or is an ordinary lines in $\mathbf{R}^n$ together with the single ideal point associated to the family of lines in $\mathbf{R}^n$ parallel to that ordinary line.

Define a map

$$\begin{aligned}\pi : \ \mathbf{P}^n - \text{ideal points} &\to \qquad \mathbf{R}^n.\\ [x_1, x_2, \ldots, x_{n+1}] &\to \left(\frac{x_1}{x_{n+1}}, \frac{x_2}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right)\end{aligned}$$

**Definition.** The map $\pi$ is called the *standard projection of $\boldsymbol{P}^n$ onto $\boldsymbol{R}^n$*.

Note that the map $\pi$ is not defined on all of $\mathbf{P}^n$. It corresponds to finding the intersection of the line through the origin and $(x_1, x_2, \ldots, x_{n+1})$ in $\mathbf{R}^{n+1}$ with the plane $x_{n+1} = 1$.

Like in the projective plane, there are many ways to coordinatize the points of $\mathbf{P}^n$. One can also define the cross-ratio of four points on a projective line.

**Definition.** A *projective transformation* or *projectivity* of $\mathbf{P}^n$ is any one-to-one and onto map $T : \mathbf{P}^n \to \mathbf{P}^n$ that preserves collinearity and the cross-ratio of points.

One can prove that projective transformation of $\mathbf{P}^n$ are defined by means of homogeneous equations in $n + 1$ variables and can be described by means of $(n + 1) \times (n + 1)$ nonsingular matrices.

**Definition.** Two figures $\mathbf{F}$ and $\mathbf{F}'$ are *projectively equivalent* if there is an projective transformation T with $T(\mathbf{F}) = \mathbf{F}'$.

**3.5.1. Theorem.** (The Fundamental Theorem of Projective Geometry for $\mathbf{P}^n$) Given two sets $\{\mathbf{P}_i\}$ and $\{\mathbf{P}_i'\}$ of n + 2 points in $\mathbf{P}^n$ with the property that no n + 1 points of either set lies in a hyperplane, then there is a unique projective transformation T of $\mathbf{P}^n$ that sends $\mathbf{P}_i$ to $\mathbf{P}_i'$.

**Proof.** The proof is analogous to what we did for $\mathbf{P}^2$.

**3.5.2. Corollary.** Given two k-dimensional projective planes $\mathbf{X}$ and $\mathbf{Y}$ in $\mathbf{P}^n$, there is a projective transformation T of $\mathbf{P}^n$ that sends $\mathbf{X}$ onto $\mathbf{Y}$.

The reader should remember two fundamental ideas: One is that, as was pointed out earlier, whenever one works with homogeneous coordinates one is really dealing with projective space, whether one is consciously thinking about that or not, because those coordinates are the natural coordinates for projective space. Second, if one has to deal with (projective) transformations of $\mathbf{R}^n$, then it is often simpler to translate the problem into one involving $\mathbf{P}^n$, to solve the corresponding problem in that space, and finally to map the answer back down to $\mathbf{R}^n$ (equivalently, solve the problem using homogeneous coordinates first). This idea can be expressed very compactly by the commutative diagram

$$\begin{array}{ccc} \mathbf{P}^n & \xrightarrow{F} & \mathbf{P}^n \\ \iota\uparrow\cup & & \cup\downarrow\pi \\ \mathbf{R}^n & \xrightarrow[f]{} & \mathbf{R}^n \end{array}$$

If one needs to deal with a transformation f, then deal with its lift F to $\mathbf{P}^n$ instead, where F is defined by the equation $f = \pi F \iota$ with $\iota$ and $\pi$ the standard inclusion and projection, respectively. (Recall that $\pi$ is actually not defined on all of $\mathbf{P}^n$.)

## 3.5.1 Homogeneous Coordinates and Maps in 3-Space

The homogeneous coordinates of a point (x,y,z) in $\mathbf{R}^3$ are a 4-tuple

$$(X, Y, Z, W), \quad \text{where } W \neq 0 \text{ and } x = X/W, \quad y = Y/W, \quad \text{and } z = Z/W.$$

Every projective transformation of $\mathbf{R}^3$ is an affine transformation or a composite of an affine transformation and a single perspectivity. The affine part can in turn be decomposed into a composition of translations, rotations, reflections, shearings, or local scaling. Using homogeneous coordinates, we can express such a map either as a product of $4 \times 4$ matrices that correspond to the maps in that composition or via a single $4 \times 4$ matrix whose parts can be described as shown below:

$$
\begin{array}{c}
\text{rotation, reflection} \\
\text{shearing, local scaling}
\end{array}
\qquad\qquad
\begin{array}{c}
\text{perspective}
\end{array}
$$

$$
\left[\begin{array}{ccc|c}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24} \\
a_{31} & a_{32} & a_{33} & a_{34} \\
\hline
a_{41} & a_{42} & a_{43} & a_{44}
\end{array}\right]
$$

$$
\text{translation} \qquad\qquad \text{global scaling}
$$

The above matrix for a projective transformation is not unique. One can always multiply each entry by a fixed nonzero constant and the resultant matrix will also represent the same transformation.

As before, there is one matrix that is of particular interest. Consider the matrix

$$
M_a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

This matrix represents a map $C_a : \mathbf{R}^3 \to \mathbf{R}^2$ that is the central projection of $\mathbf{R}^3$ onto the x-y plane from the point $(0,0,-1/a)$ on the z-axis (the analog of the planar map $C_a$ in Section 3.4.3). To see this, note that

$$
[x\ y\ z\ 1]M_a = [x\ y\ 0\ \ az+1].
$$

In other words,

$$
C_a(x,y,z) = \left(\frac{x}{az+1}, \frac{y}{az+1}\right),
$$

which agrees with what this particular central projection should do.

**3.5.1.1. Example.** Find the central projection S of $\mathbf{R}^3$ from the origin onto the plane $\mathbf{X}$ defined by the equation $2x - y + z = 6$.

**Solution.** We shall solve this problem in three different ways. The first and most trivial solution is simply to find the intersection of lines through the origin with the plane $\mathbf{X}$. Now the parametric equation of the line $\mathbf{L}$ through the origin and the point $(x,y,z)$ are

$$x' = xt,$$
$$y' = yt,$$
$$z' = zt.$$

But $(x',y',z')$ belongs to **X** if and only if

$$2xt - yt + zt = 6,$$

that is,

$$t = \frac{6}{2x - y + z}.$$

In other words, S is defined by the equations

$$x' = \frac{6x}{2x - y + z}, \quad y' = \frac{6y}{2x - y + z}, \quad z' = \frac{6z}{2x - y + z}.$$

A second way to solve the problem is to transform the problem into simpler ones. The y-, x-, and z intercepts of **X** are the points **A**(0,–6,0), **B**(3,0,0), and **C**(0,0,6), respectively. See Figure 3.16. To find the map S, we shall reduce this problem to one we can handle by describing S as a composition of three maps for which we already know how to derive the equations. First of all, let T be the linear transformation of $\mathbf{R}^3$, which sends **A** to **A′**(0,–6,6), **B** to **B′**(3,0,6), and **C** to **C′** = **C**. It is easy to see that T sends the plane **X** to the plane defined by z = 6. Next, let R be the translation that translates this plane to the x-y coordinate plane. Then RT sends the origin to (0,0,–6). It follows that $S = T^{-1}R^{-1}C_{1/6}RT$. The maps have the following matrices (with respect to homogeneous coordinates):

$$T \leftrightarrow \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad T^{-1} \leftrightarrow \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

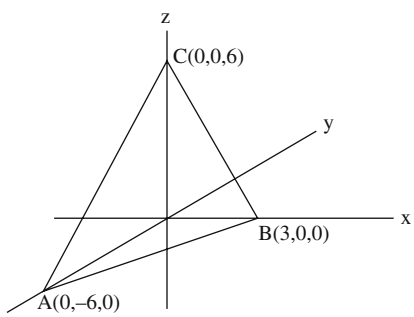



**Figure 3.16.** The plane in Example 3.5.1.1.

$$R \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -6 & 1 \end{pmatrix} \qquad R^{-1} \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 6 & 1 \end{pmatrix} \qquad C_{1/6} \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Multiplying these matrices together (in the opposite order from the maps) we get that S has matrix

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & -\frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.31}$$

This means that the homogeneous equations for S are

$$\begin{aligned} x' &= x \\ y' &= y \\ z' &= z \\ w' &= \frac{1}{3}x - \frac{1}{6}y + \frac{1}{6}z \end{aligned}$$

Dividing through by w′ leads to the same nonhomogeneous equations for S as before. Note that these equations give the right results for the points **A**, **B**, and **C**.

A third solution to the problem uses strictly rigid motions and frames. (The transformation T above is not a rigid motion.) Our first goal is to find an orthonormal basis $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ for 3-space so that the first two vectors are an orthonormal basis for **X**. To find $\mathbf{u}_1$ and $\mathbf{u}_2$ we use the Gram-Schmidt algorithm on any basis of **X**. Using the basis **AB** and **AC** leads to

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}(1,2,0) \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sqrt{30}}(-2,1,5).$$

Since the equation for **X** implies that (2,–1,1) is a normal vector for **X** we can let

$$\mathbf{u}_3 = \frac{1}{\sqrt{6}}(2,-1,1).$$

Let F be the frame $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$. Then $F^{-1}$ transforms a point (x,y,z) into F-coordinates (x′,y′,z′). In this coordinate system the plane **X** becomes the plane **X′** defined by the equation $z' = \sqrt{6}$. (**X** was a distance $\sqrt{6}$ from the origin.) In other words, if R is the translation with translation vector $(0,0,-\sqrt{6})$, then

$$S = FR^{-1}C_{1/\sqrt{6}}RF^{-1}.$$

In homogeneous coordinates the matrices for $F^{-1}$ and F are

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} & -\frac{1}{\sqrt{6}} & 0 \\ 0 & \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} & 0 & 0 \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{30}} & \frac{5}{\sqrt{30}} & 0 \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

respectively. Since the homogeneous matrix for $R^{-1}C_{1/\sqrt{6}}R$ is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{\sqrt{6}} \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

a simple multiplication of matrices gives us the same homogeneous matrix for S as in (3.31). This shows that our new answer to the problem agrees with the two previous answers.

If all one wants is the equations for S in Cartesian coordinates, then clearly the first solution to this problem is the simplest. The second solution was given mainly to emphasize the fact that **any** transformation can be used in solving such problems as long as they preserve the relevant aspects of the problem, in this case lines and intersections of lines. Affine maps clearly do that. However, sometimes, as in computer graphics applications, one is really after the equations for the central projection in frame F coordinates. This amounts to finding the equations for the transformation $R^{-1}C_{1/\sqrt{6}}RF^{-1}$ above. In that case the approach taken in the third solution needs to be followed and this is what one must fully understand since the first two approaches are not relevant here.

**3.5.1.2. Example.** Use frames to find the central projection C of the x-y plane onto the line **L** defined by the equation $x - y = 2$ from the point $\mathbf{p} = (5,1)$.

**Solution.** See Figure 3.17. This is the same problem as in Example 3.4.3.1. We want a frame $F = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{p})$ so that $\mathbf{u}_1$ is a basis for **L** and $\mathbf{u}_2$ is normal to **L** and points from **p** "to" **L**. A natural choice is

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,1) \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}(-1,1).$$

The direction of $\mathbf{u}_1$ does not matter here. Since the distance from **p** to **L** is $\sqrt{2}$, if we let T be the translation with translation vector $-\sqrt{2}$, then

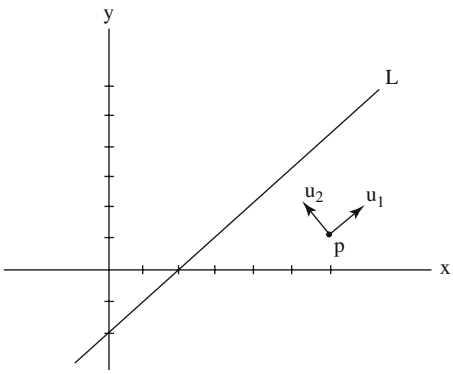



**Figure 3.17.** Using frames for central projections.

$$C = FT^{-1}C_{1/\sqrt{2}}TF^{-1}.$$

It is easy to check that our answer agrees with the one in Example 3.4.3.1.

## 3.6 Conic Sections

This section defines the classic conic sections or conics and discusses their geometry. We analyze the equations that define them and derive the well-known formulas that are used to classify them. An excellent general reference on the geometry of these curves is [HilC99]. All our points and sets in this section are assumed to lie in $\mathbf{R}^3$.

**Definition.** Let $\mathbf{X}$ be a plane, $\mathbf{C}$ a circle in $\mathbf{X}$, and $\mathbf{p}$ a point not contained in $\mathbf{X}$. The union of all the lines through $\mathbf{p}$ and a point of the circle $\mathbf{C}$ is called a (*circular*) *cone*. The point $\mathbf{p}$ is called the *vertex* of the cone. The line through $\mathbf{p}$ and the center of the circle $\mathbf{C}$ is called the *axis* of the cone. If the axis of the cone is orthogonal to $\mathbf{X}$, then the cone is called a *right circular cone*; otherwise it is called an *oblique circular cone*.

**Note.** We are using very common terminology here, but we should point out that the term "cone" is being used here in the sense of a "cone of lines." In other contexts like topology a more accurate term for what we are calling a "cone" would be "double cone." See Section 5.4.

**Definition.** A *conic section* is any set of points obtained as the intersection of a circular cone and a plane. It the plane passes through the vertex of the cone, then the conic section is called *degenerate*; if not, the conic section is called *nondegenerate*.

See Figure 3.18 for examples of some conic sections. The next theorem is helpful in deriving an analytic version of the definition of a conic section.

**3.6.1. Theorem.** Every nondegenerate conic section is a set of points in a plane that is either a circle with positive radius or a set with the property that the ratio e of the

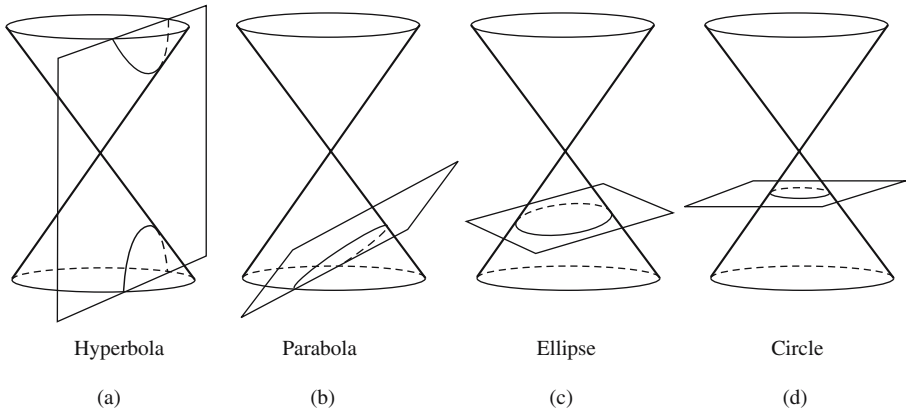



**Figure 3.18.** Conic sections.

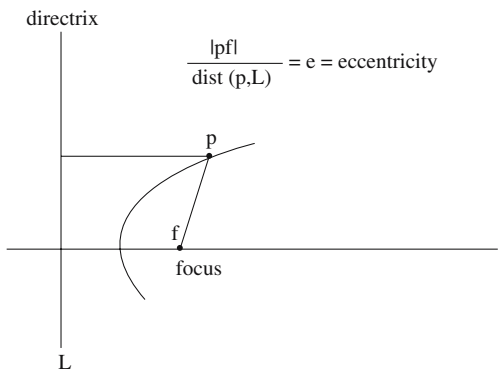



**Figure 3.19.** Some notation for conics.

distance of each of its points from a fixed point **p** to the distance of the point from a fixed line **L** not passing through **p** is constant.

**Proof.** See [Eise39].

**Definition.** Using the notation in Theorem 3.6.1 with respect to some noncircular, nondegenerate conic section **C**, the fixed point **p** is called the *focus*, the fixed line **L** is called the *directrix*, and the constant e is called the *eccentricity* of **C**. The conic section is called an *ellipse*, *parabola*, or *hyperbola* depending on whether $e < 1$, $e = 1$, or $e > 1$, respectively. Ellipses and hyperbolas are often called *central conic sections*. Define the *focus* of a circle to be its center and its *eccentricity* to be 0. (We can make the definition of a circle match that of the other conic sections completely by thinking of the *directrix* of a circle as a line at infinity and the limiting case for an ellipse where we let its directrix move further and further away from the focus.)

See Figure 3.19. Conic sections can also be constructed via string constructions. See [HilC99]. For example, we can trace out an ellipse by tying the two ends of a string

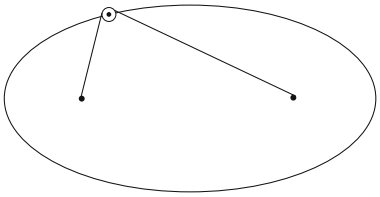


**Figure 3.20.** Pencil and string construction for an ellipse.

of fixed length to the foci, stretching the string tightly with a pencil, and drawing all the points that can be reached by the pencil in that way. See Figure 3.20. The next theorem expresses this geometric characterization and related ones for the hyperbola and parabola mathematically.

### 3.6.2 Theorem

(1) An ellipse can be defined as a set of points such that the sum of their distances to two fixed points is constant. The two points are the foci of the ellipse.
(2) A hyperbola can be defined as a set of points such that the difference of their distances to two fixed points is constant. The two points are the foci of the hyperbola.
(3) A parabola can be defined as a set of points such that the sum of their distances to a fixed point and a fixed line is constant.

**Proof.** See [Eise39] or [Full73].

**Definition.** Two central conic sections in the plane are said to be *confocal* if they have the same foci.

One can show that the family of confocal ellipses for two fixed foci covers the plane with each point in the plane belonging to a unique ellipse in the family. A similar fact holds for confocal hyperbolas. See [HilC99]. Furthermore, if an ellipse and a hyperbola have the same foci, then the two curves intersect orthogonally.

**Definition.** The points of a conic section where it intersects the line through the focus that is orthogonal to the directrix are called the *vertices* of that conic section.

It is easy to show that parabolas have one vertex and ellipses and hyperbolas have two vertices. Now let $\mathbf{C}$ be a conic section and let $\mathbf{L}_1$ be the line through its focus $\mathbf{p}$ that is orthogonal to the directrix. Define a point $\mathbf{O}$ as follows: If $\mathbf{C}$ is a parabola, then $\mathbf{O}$ is its vertex, otherwise, we have two vertices and $\mathbf{O}$ is the midpoint of the segment that has them as end points. Let $\mathbf{L}_2$ be the line through $\mathbf{O}$ orthogonal to $\mathbf{L}_1$. The orthogonal lines $\mathbf{L}_1$ and $\mathbf{L}_2$ determine a natural coordinate system for our conic section. Let $\mathbf{u}_1 = \mathbf{Op}/|\mathbf{Op}|$, and let $\mathbf{u}_2$ be a unit direction vector for $\mathbf{L}_2$.

**Definition.** The coordinate system for the plane containing the conic section $\mathbf{C}$ determined by the frame $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{O})$ is called the *natural coordinate system* for the conic

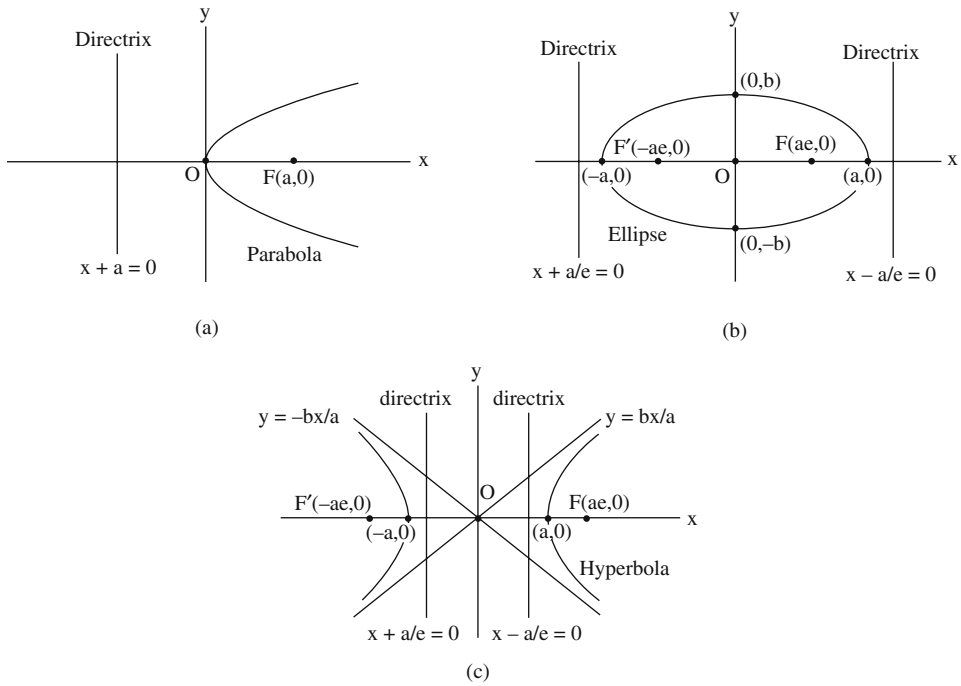



**Figure 3.21.** The natural coordinate systems for conic sections.

section. If **C** is a parabola, then $\mathbf{L}_1$ is called its *axis*. If **C** is an ellipse, then $\mathbf{L}_1$ and $\mathbf{L}_2$ will be called the *major* and *minor (coordinate) axis* of **C**, respectively. If **C** is a hyperbola, then $\mathbf{L}_1$ and $\mathbf{L}_2$ will be called the *transverse* and *conjugate (coordinate) axis* of **C**, respectively. In the case of either an ellipse or hyperbola the point **O** is called its *center*.

Figure 3.21 shows the coordinate systems and coordinate axes for parabolas, ellipses, and hyperbolas. Ellipses and hyperbolas are called central conic sections because they have a center. One can show that the conic sections are symmetric about their axes, meaning that if a point belongs to them, then the reflected point about their axes will belong to them also.

In the case of the plane $\mathbf{R}^2$, we can use Theorem 3.6.1 and write out the constraints on the distances of points from the focus and directrix in terms of an equation. It is easily seen to be a quadratic equation. Let us look at the equations for some special well-known cases in their natural coordinate system.

**The parabola:** $y^2 = 4ax$ (3.32)

Focus: $(a, 0)$, $a > 0$
Directrix: $x = -a$
Eccentricity: $e = 1$

**The ellipse:** $$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \quad a \geq b > 0 \tag{3.33}$$

Focus: (c, 0)

Directrix: $x = \dfrac{a^2}{c}$ (if $c > 0$)

Eccentricity: $e = \dfrac{c}{a}$, where $c \geq 0$ and $c^2 = a^2 - b^2$

The segments [(–a,0),(a,0)] and [(0,–b),(0,b)] are also sometimes called the *major* and *minor axis* of the ellipse, respectively, or simply the *principal axes*.

**The hyperbola:** $$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1, \quad a,b > 0 \tag{3.34}$$

Focus: (c, 0)

Directrix: $x = \dfrac{a^2}{c}$

Eccentricity: $e = \dfrac{c}{a}$, where $c^2 = a^2 + b^2$

The segments [(–a,0),(a,0)] and [(0,–b),(0,b)] are also sometimes called the *transverse* and *conjugate axis* of the hyperbola, respectively. The *asymptotes* of the hyperbola are the lines $y = \pm bx/a$.

Because of the symmetry present in the case of a hyperbola and ellipse, they actually have two foci and directrices. The second of the pair is obtained by reflecting the ones given above about the y-axis.

The discussion above shows that conic sections are solutions to quadratic equations. To connect the two we start from the other direction.

**Definition.** An (*affine*) *conic* is any subset of $\mathbf{R}^2$ defined by an equation of the form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

where $(a,b,c) \neq (0,0,0)$.

Our object is to show that the terms "conic" and "conic section" refer essentially to the same geometric spaces. To that end it is convenient to rewrite the equation defining a conic in the form

$$ax^2 + 2hxy + by^2 + 2fx + 2gy + c = 0. \tag{3.35}$$

Although this may look a little artificial right now, it prevents a factor of 2 from entering equations at a later stage and so we shall use this form of the equation from now on. In any case, to understand the solution set of our quadratic equation, the idea will be to transform it into a simpler one. We shall show that in the nondegenerate cases a simple change in coordinate systems will change the general quadratic equation (3.35) into one of the equations (3.32), (3.33), or (3.34). The main problem is finding a change in coordinates that will eliminate the xy term.

We start off by describing two straightforward simple-minded but ad hoc approaches to solve our problem. The second is actually good enough to give us the formulas that one uses to convert (3.35) to one of the standard forms. However, there are also some well-defined invariants associated to (3.35), which can tell us right away what sort of curve one has without actually transforming the equation. To be able to prove that these invariants work as specified is what motivates us to present a third, more elegant approach to analyzing (3.35) using homogeneous coordinates and the theory of quadratic forms.

The first simple-minded approach uses a form of "completing the square." Probably the most straightforward way to do this is to rewrite the equation as

$$\left[a_1x+\frac{h}{a_1}y\right]^2+\frac{ab-h^2}{a}y^2+2fx+2gy+c=0$$

where $a_1$ satisfier $a = a_1^2$. The substitution

$$\begin{aligned} x' &= a_1x-\frac{h}{a_1}y \\ y' &= \qquad\quad y \end{aligned}$$

will then produce an equation in $x'$ and $y'$ that has no $x'y'$ term. Although this substitution is satisfactory in some applications and produces simple formulas, it has the disadvantage that the linear transformation of coordinates to which it corresponds is not a rigid motion and may deform shapes in undesirable ways. Thus, since the xy term arises from a rotation of the axes of the conic, a second and "better" way to eliminate this term is via a rotation about the origin. Consider the substitution

$$\begin{aligned} x &= x'\cos\theta-y'\sin\theta \\ y &= y'\sin\theta+y'\cos\theta \end{aligned}$$

which corresponds to rotating the conic about the origin through an angle $-\theta$. This substitution will transform equation (3.35) into an equation in $x'$ and $y'$ for which the coefficient of the $x'y'$ term is

$$2(b-a)\sin\theta\cos\theta+2h\left(\cos^2\theta-\sin^2\theta\right).$$

Setting this expression to zero and using some simple trigonometric identities gives the equation

$$(b-a)\sin 2\theta + 2h\cos 2\theta = 0.$$

If a = b, then $\cos 2\theta = 0$, which means that $\theta$ is ±45 degrees. If $a \neq b$, then the solution can be written either as

$$\tan 2\theta = \frac{2h}{a-b} \tag{3.36a}$$

or

$$\tan\theta = \frac{b-a \pm \sqrt{(b-a)^2 + 4h^2}}{2h}. \tag{3.36b}$$

Note that if $\theta_1$ and $\theta_2$ are angles satisfying (3.36b) where we use the + and − sign, respectively, in the formula, then

$$\tan\theta_1 \tan\theta_2 = -1,$$

which shows that the angles differ by 90 degrees. We shall see that this basically affirms that the "axes" of the conic are perpendicular. In any case, there is an angle $\theta$ which will eliminate the $x'y'$ cross-term in equation (3.35). Thus, we end up with an equation of the form

$$a'x'^2 + b'y'^2 + 2f'x' + 2g'y' + c' = 0. \tag{3.37}$$

The rest of the steps involved in analyzing equation (3.35) are very simple. Equation (3.37) still has some linear terms that need to be eliminated if the corresponding quadratic term is present. This is done by completing the square in the standard way. For example, if $a' \neq 0$, then make the substitution

$$x' = x'' + \frac{g'}{a'}.$$

From this one sees that, in the nondegenerate case, equation (3.38) splits into two cases. If $a'b' = 0$, then (3.35) represents a parabola; otherwise, (3.35) represents an ellipse or hyperbola depending on whether the sign of $a'b'$ is positive or negative.

In summary, we have shown how some simple manipulations of equation (3.35) enable us to determine the geometry of the solution set. However, we have glossed over some degenerate cases. For example, a degenerate case of equation (3.35) occurs when it factors into two linear terms, that is, that it can be written in the form

$$(a_1x + b_1y + c_1)(a_2x + b_2y + c_2) = 0. \tag{3.38}$$

In this case the equation represents a pair of lines. To be able to detect the special cases, and all the other cases for that matter, in a nice simple way, we need to make use of some more powerful tools. In particular, we need to switch to homogeneous coordinates and projective space.

**Definition.** A *conic in the projective plane*, or a *projective conic*, is a set of points in $\mathbf{P}^2$ whose homogeneous coordinates (x,y,z) satisfy a quadratic equation of the form

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz = 0, \tag{3.39}$$

where $(a,b,d) \neq (0,0,0)$.

First, projective conics are well defined in the sense that if one homogeneous coordinate representative of a point in $\mathbf{P}^2$ satisfies equation (3.39), then they all do. Second, every conic in the plane defines a unique projective conic because, switching to homogeneous coordinates, equation (3.35) changes into

$$ax^2 + 2hxy + by^2 + 2fxz + 2gyz + cz^2 = 0. \tag{3.40}$$

Third, the condition that at least one of the values a, b, and d be nonzero is simply the analog of the condition on the coefficients of an affine conic. It is an extremely artificial condition at the projective level. The fact is that when someone says "conic" one thinks of ellipses, hyperbolas, and parabolas. We were trying to exclude grossly degenerate cases of equation (3.39) such as 0 = 0, which would have all of $\mathbf{P}^2$ as its solution. The projective conics we are after are the "nondegenerate" ones that are defined below. Our condition on the coefficients happens to capture (with foresight) the equation form of the conic sections. Having said this, the fact that one can give a complete analysis of the solutions to the general quadratic equation (3.39) is interesting on its own, independent of any relation to conic sections. This equation-solving mindset is the context in which we will be working right now.

Now equation (3.40) is also the convenient form for projective conics. Consider the quadratic form

$$q_1(x,y,z) = ax^2 + 2hxy + by^2 + 2fxz + 2gyz + cz^2$$

and its matrix

$$A = \begin{pmatrix} a & h & f \\ h & b & g \\ f & g & c \end{pmatrix}. \tag{3.41}$$

Note that equation (3.40) can be rewritten in matrix form as

$$(x,y,z)A(x,y,z)^T = 0. \tag{3.42}$$

Conversely, equation (3.42) defines a projective conic for any symmetric $3 \times 3$ matrix A satisfying $(a,b,h) \neq (0,0,0)$. The next theorem shows that projective conics are actually very easy to describe if one chooses the correct coordinate system.

**3.6.3. Theorem.** We can coordinatize projective space $\mathbf{P}^2$ in such a way that in the new coordinate system the equation of the projective conic defined by equation (3.40) has one of the following forms:

$$z^2 = 0 \tag{3.43a}$$

$$x^2 \pm z^2 = 0 \tag{3.43b}$$

$$x^2 + y^2 \pm z^2 = 0 \tag{3.43c}$$

**Proof.** By Theorem 1.9.11, any symmetric matrix, the matrix A in (3.41) in particular, is congruent to a diagonal matrix with ±1 or 0 along the diagonal. This clearly implies the result since A is not the zero matrix. Note that since we are working in projective space here the change of coordinates transformations that produce equations (3.43) are **projective** and not affine transformation in general.

Translating things back to Euclidean space, it is easy to see that equations (3.43a) and (3.43b) correspond to cases where the solution set to (3.35) is either empty or consists of lines. Equations (3.43c) is the case where A is nonsingular.

**Definition.** The affine conic defined by equation (3.35) or the projective conic defined by equation (3.40) is said to be *nondegenerate* if the matrix A in (3.41) is nonsingular; otherwise it is said to be *degenerate*.

**Note.** The definition of a nondegenerate conic has the advantage of simplicity but has a perhaps undesirable aspect to it, at least at first glance. If A is nonsingular, then one possibility for equation (3.43c) is

$$x^2 + y^2 + z^2 = 0 \quad \left(\text{or} \quad x^2 + y^2 + 1 = 0 \quad \text{using Cartesian coordinates}\right).$$

This equation has no **real** nonzero solutions. Therefore, a "nondegenerate" conic could be the empty set. For that reason, some authors add the condition that a conic be nonempty before calling it nondegenerate. On the other hand, we would get a nonempty set if we were to allow complex numbers and we were talking about conics in the complex plane. See Section 10.2.

We have just seen that matrix A in (3.41) determines one important invariant for conics, but there is another. Consider the quadratic form that is the homogeneous part of equation (3.35), namely,

$$q_2(x, y) = ax^2 + 2hxy + by^2. \tag{3.44}$$

Let

$$B = \begin{pmatrix} a & h \\ h & b \end{pmatrix} \tag{3.45}$$

be the matrix associated to $q_2$. By Theorem 1.9.10 there is a change of basis that will diagonalize B, that is, in the new coordinate system $q_2$ will have the form

$$q_2(x',y') = a'x'^2 + b'y'^2.$$

The numbers a′ and b′ are just the eigenvalues of the transformation associated to B and are the roots of the characteristic polynomial for B. Since the change corresponds to a **linear** change of variables, in this new coordinate system equation (3.35) will have been transformed into an equation of the form (3.37). We now have all the pieces of the puzzle.

**3.6.4. Theorem.** Define numbers Δ, D, and I for equation (3.35) by

$$\Delta = \begin{vmatrix} a & h & f \\ h & b & g \\ f & g & c \end{vmatrix}, \quad D = \begin{vmatrix} a & h \\ h & b \end{vmatrix} = ab - h^2, \quad \text{and} \quad I = a + b.$$

(1) The quantities Δ, D, and I are invariant under a change of coordinates via a rigid motion (translation or rotation).
(2) If $\Delta \neq 0$, then equation (3.35) defines a nondegenerate conic. More precisely,

(a) $D > 0$: We have an ellipse if $I\Delta < 0$ and the empty set otherwise. (Note that since a and b have the same sign in this case, the sign of IΔ is the same as the sign of bΔ or aΔ.)
(b) $D < 0$: We have a hyperbola.
(c) $D = 0$: We have a parabola.

(3) If $\Delta = 0$, then equation (3.35) factors into two factors of degree one (the converse is also true) and defines the empty set, a point, or a pair of lines. The pair of lines may be parallel, intersecting, or coincident. More precisely,

(a) $D > 0$: We get a single point.
(b) $D < 0$: We get two intersecting lines.
(c) $D = 0$:

$b \neq 0$: There are three cases depending on $E = g^2 - bc$.

$E > 0$: We get two parallel lines.
$E < 0$: We get the empty set.
$E = 0$: We get a single line.

$b = 0$: Then $h = 0$. There are three cases depending on $F = f^2 - ac$.

$F > 0$: We get two parallel lines.
$F < 0$: We get the empty set.
$F = 0$: We get a single line.

**Proof.** Part (1) follows from properties of the determinant and the trace function (I is the trace of the matrix B). The main ideas behind the proof of (2) have been sketched above. The condition on the product IΔ in (2)(a) is equivalent to saying that I and Δ have opposite signs. The effect that the signs of I and Δ have on the conic is best seen

by looking at the matrices after they have been diagonalized. The details of the rest of the proof are lengthy and messy but not hard. They basically only involve rewriting equation (3.35) and solving for various conditions. See [Eise39].

Theorem 3.6.4 tells us the conic that equation (3.35) represents but does not directly tell us the transformation that transforms it to our standard equations. The basic steps in finding this transformation were sketched earlier. We summarize the results along with some additional details below. They are divided into three cases that subdivide into the subcases in Theorem 3.6.4. Again see [Eise39]. For a slightly different approach, see [RogA90].

Let **C** be the conic defined by equation (3.35). Our object is to find a rigid motion M so hat M(**C**) is defined by the standard equation for a conic. The transformation M will have the form RT, where T is a translation and R is a rotation about the origin.

**Case 1:** $D \neq 0$ (the central conics if $\Delta \neq 0$)

The "center" of the conic is at

$$x_0 = \frac{hg - bf}{D}, \quad y_0 = \frac{hf - ag}{D}. \tag{3.46}$$

Let T be the translation that sends $(x_0,y_0)$ to the origin. The conic $\mathbf{C}' = T(\mathbf{C})$ will have its center at the origin. It follows that replacing x by $x + x_0$ and y by $y + y_0$ will eliminate the linear terms in (3.35) and give us an equation for $\mathbf{C}'$ of the form

$$ax^2 + 2hxy + by^2 + \text{constant} = 0. \tag{3.47}$$

Let $\theta$ be an angle defined by equations (3.36) and let R be the rotation about the origin through the angle $-\theta$. This will rotate an axis of the conic into the x-axis. The equation for $\mathbf{C}'' = R(\mathbf{C}')$ is obtained by replacing x by $x \cos\theta - y \sin\theta$ and y by $x \sin\theta + y \cos\theta$ in equation (3.47). We will get

$$a'x'^2 + b'y'^2 + \frac{\Delta}{D} = 0. \tag{3.48}$$

In fact, a′ and b′ are the roots of

$$x^2 - Ix + D = 0, \tag{3.49}$$

that is,

$$a' = \frac{a + b + \sqrt{(a-b)^2 + 4h^2}}{2} \quad \text{and} \quad b' = \frac{a + b - \sqrt{(a-b)^2 + 4h^2}}{2}. \tag{3.50}$$

Equation (3.49) is actually the characteristic equation for the matrix B in (3.45) (Theorem C.4.8(3)), so that a′ and b′ are the eigenvalues of B. In other words, instead of dealing with angles directly, find an orthonormal pair $\mathbf{u}_1$ and $\mathbf{u}_2$ of eigenvectors for B. Let

$$(x\ y) = (x'\ y')\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \tag{3.51}$$

and substitute the x and y that one gets from equation (3.51) into equation (3.47). This will also produce equation (3.48).

**Case 2:** D = 0 and $\Delta \neq 0$ (parabola)

One can show that the vertex of the parabola is at $(x_0,y_0)$, where $x_0$ and $y_0$ are the solution to the equations

$$ax_0 + hy_0 + \frac{af + hg}{I} = 0 \tag{3.52}$$

and

$$\left(f + \frac{fb - hg}{I}\right)x_0 + \left(g + \frac{ag - hf}{I}\right)y_0 + c = 0. \tag{3.53}$$

Furthermore,

$$(a + b)(hx + by) + fh + gb = 0 \tag{3.54}$$

is the equation of the axis of the parabola. Therefore, translate the origin to the vertex of the parabola and then rotate about that point through an angle $-\theta$, where

$$\tan\theta = -\frac{h}{b} = -\frac{a}{h}. \tag{3.55}$$

These transformations will change equation (3.35) into

$$y'^2 = 2px', \tag{3.56a}$$

where

$$p = \frac{ag - hf}{I\sqrt{a^2 + h^2}}. \tag{3.56b}$$

**Case 3:** D = 0 and $\Delta = 0$ (degenerate cases)

If $b \neq 0$, then equation (3.35) reduces to

$$y = \frac{-(hx + g) \pm \sqrt{g^2 - bc}}{b}. \tag{3.57}$$

If b = 0 (and also h = 0), then we get

$$x = \frac{-f \pm \sqrt{f^2 - ac}}{a}. \tag{3.58}$$

Alternatively, rotate about the origin through the angle $-\theta$, where $\theta$ is defined by equation (3.55). This changes equation (3.35) into

$$Iy'^2 + 2\frac{af + hg}{\sqrt{a^2 + b^2}}y' + c = 0. \tag{3.59}$$

If there are real roots to this equation, then we get a factorization

$$(y' - y_1)(y' - y_2) = 0,$$

which corresponds to one or two straight lines.

**3.6.5. Example.** To transform

$$52x^2 - 72xy + 73y^2 + 8x - 294y + 333 = 0 \tag{3.60}$$

into standard form.

**Solution.** We have

$$A = \begin{pmatrix} 52 & -36 & 4 \\ -36 & 73 & -147 \\ 4 & -147 & 333 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 52 & -36 \\ -36 & 73 \end{pmatrix}.$$

Therefore, $I = 125$, $\Delta = \det(A) = -250000$, and $D = \det(B) = 2500$ and we fall into Case 1. In fact, $D > 0$, $\Delta \neq 0$, and $I\Delta < 0$ means that we have an ellipse. After translating the center of the ellipse, defined by equation (3.46), to the origin, we can finish the reduction of equation (3.60) in two ways: we can rotate through the angle specified by equations (3.36), or we could use the eigenvector approach indicated by equation (3.51). Of course, we could also just simply use equations (3.48) and (3.50), but, although this may seem simpler, the advantage of the other two methods is that they also give us the coordinate transformation that transforms the standard coordinate system into the one in which the curve has the standard form. One often needs to know this transformation.

We start with the angle approach. Using equation (3.46), the center $(x_0, y_0)$ of the curve turns out to be (2,3). Substituting $x + 2$ and $y + 3$ for x and y in equation (3.60) gives

$$52x^2 - 72xy + 73y^2 - 100 = 0. \tag{3.61}$$

We now want to rotate the coordinate axes through an angle $-\theta$, where, using formula (3.36b),

$$\tan\theta = -\frac{4}{3} \quad \text{or} \quad \frac{3}{4}.$$

We arbitrarily choose 3/4, so that

$$\cos\theta = \frac{4}{5} \quad \text{and} \quad \sin\theta = \frac{3}{5},$$

and the equation for the rotation R about the origin through the angle $-\theta$ is

$$x' = \frac{4}{5}x + \frac{3}{5}y$$

$$y' = -\frac{3}{5}x + \frac{4}{5}y.$$

Let **C** be the curve defined by (3.60). Then the curve $\mathbf{C}'' = \mathrm{RT}(\mathbf{C})$ has equation

$$x^2 + 4y^2 - 4 = 0 \tag{3.62}$$

and we are done.

Next, we use the eigenvalue approach. The eigenvalues are the roots of equation (3.49), which reduces to

$$x^2 - 125x + 2500 = (x-25)(x-100) = 0$$

in this case. To solve for the eigenvectors for eigenvalues 25 and 100 we must solve

$$(x\ y)\left(25I^2 - B\right) = 0 \quad \text{and} \quad (x\ y)\left(100I^2 - B\right) = 0,$$

that is,

$$(x\ y)\begin{pmatrix} -27 & 36 \\ 36 & -48 \end{pmatrix} = (0\ 0) \quad \text{and} \quad (x\ y)\begin{pmatrix} 48 & 36 \\ 36 & 27 \end{pmatrix} = (0\ 0),$$

respectively. The solutions for the first equation are $x = (4/3)y$, so that $\mathbf{u}_1 = (4/5,3/5)$ is a unit eigenvector for eigenvalue 25. The solutions for the second equation are $x = (-3/4)y$, so that $\mathbf{u}_2 = (-3/5,4/5)$ is a unit eigenvector for eigenvalue 100. The substitution specified by equation (3.51) would then again give us equation (3.62). It corresponds to the same rotation R described above.

One point that one needs to be aware of when using the eigenvalue approach is that there is some leeway as to our choice of eigenvectors. Our only real constraint is that the orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2)$ induce the standard orientation of the plane because we want a rigid motion, specifically, a rotation. On the other hand, $(\mathbf{u}_2,-\mathbf{u}_1)$ would have been a legitimate alternative choice. This would have reduced our conic equation to

$$4x^2 + y^2 - 4 = 0.$$

But then, there are always basically two standard forms to which a general conic equation can be reduced. Which one we get depends on our choice of which axis we call the x- and y-axis.

**3.6.6. Example.** To transform

$$16x^2 + 24xy + 9y^2 - 100 = 0 \tag{3.63}$$

into standard form.

**Solution.** We have

$$A = \begin{pmatrix} 16 & 12 & 0 \\ 12 & 9 & 0 \\ 0 & 0 & -100 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 16 & 12 \\ 12 & 9 \end{pmatrix}.$$

Therefore, $I = 25$, $\Delta = \det(A) = 0$, and $D = \det(B) = 0$. By Theorem 3.5.1.3(3.c) we are dealing with two parallel lines since $E = 0^2 - 16(-100) = 1600$. Equation (3.59) implies that (3.63) can be transformed into

$$25y^2 - 100 = 0$$

via a rotation through an angle $-\theta$, where $\tan\theta = -4/3$ .

Note that equation (3.57) implies that (3.63) is equivalent to

$$y = -\frac{4}{3}x \pm \frac{10}{3},$$

which can easily be checked.

This finishes our discussion of the main results about quadratic equations in two variables. In the process we have proved the following:

**3.6.7. Theorem.** Every nonempty conic is a conic section. Conversely, if we coordinatize the intersecting plane in the definition of a conic section, then the conic section is defined by an equation of the form (3.35) in that coordinate system, that is, it is a conic.

Theorem 3.6.7 justifies the fact that the term "conic" and "conic section" are used interchangeably.

## 3.6.1 Projective Properties of Conics

This section looks at some projective properties of conics. There is an important corollary to Theorem 3.6.3.

**3.6.1.1. Theorem.** All nonempty nondegenerate (affine) conics are projectively equivalent.

**Proof.** Since the conic is nonempty and nondegenerate, Theorem 3.6.3 implies that it is projectively equivalent to a conic with equation

$$x^2 + y^2 - z^2 = 0 \quad (\text{or } x^2 + y^2 - 1 = 0 \text{ using Cartesian coordinates}).$$

This shows that every such conic is projectively equivalent to the unit circle and we are done.

**3.6.1.2. Example.** To show that the conic $y = x^2$ is projectively equivalent to the unit circle.

**Solution.** Passing to homogeneous coordinates, the conic is defined by the equation

$$x^2 - yz = 0 \tag{3.64}$$

with associated symmetric matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 0 \end{pmatrix}.$$

Using elementary matrices, we shall now show that A is congruent to a diagonal matrix. First of all, if E is the elementary matrix $E_{23}(-1)$, then

$$A_1 = EAE^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 0 \end{pmatrix}.$$

Next, let F be the elementary matrix $E_{32}(1/2)$. Then

$$A_2 = FA_1F^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{pmatrix}.$$

Finally, if G is the elementary matrix $E_{33}(2)$, then

$$A_3 = GA_2G^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

It follows that if

$$M = GFE = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix},$$

then $MAM^T$ is the diagonal matrix $A_3$, that is, our conic is projectively equivalent to the unit circle

$$x^2 + y^2 = 1,$$

and we are done.

Example 3.6.1.2 leads us to some observations about the relationship between a conic in $\mathbf{R}^2$ and the associated projective conic in $\mathbf{P}^2$. Consider the solutions to equation (3.64). One of the solutions (in $\mathbf{P}^2$) is the ideal point which has $z = 0$. Substituting this value into (3.64) defines the line $x = 0$ in $\mathbf{R}^2$. In other words, the parabola $y = x^2$ corresponds to the conic in $\mathbf{P}^2$, which contains the same real points and has one additional ideal point corresponding to the line $x = 0$. As another example, consider the hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{3.65}$$

The homogeneous equation for this conic is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - z^2 = 0. \tag{3.66}$$

The ideal points with $z = 0$ lead to the equations

$$y = \frac{b}{a}x \tag{3.67a}$$

and

$$y = -\frac{b}{a}x \tag{3.67b}$$

which define two lines in $\mathbf{R}^2$. It follows that the conic in $\mathbf{P}^2$ defined by (3.66) is topologically a circle that consists of the points defined by the real roots of equation (3.65) together with two extra (ideal) points associated to the lines in (3.67a) and (3.67b). Intuitively, if we were to walk along points (x,y) on the curve (3.65) where these points approach either $(+\infty,+\infty)$ or $(-\infty,-\infty)$ we would in either case approach the ideal point associated to the line defined by equation (3.67a). Letting x and y approach either $(-\infty,+\infty)$ or $(+\infty,-\infty)$ would bring us to the ideal point associated to the line defined by equation (3.67b).

**3.6.1.3. Theorem.** A conic can be found that passes through any given five points. It is unique if no four of these points is collinear.

**Proof.** We use homogeneous coordinates and need to show that we can always find a nondegenerate equation of the form

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz = 0,$$

which is satisfied for the points. One such equation is

$$\begin{vmatrix} x^2 & y^2 & z^2 & xy & xz & yz \\ x_1^2 & y_1^2 & z_1^2 & x_1y_1 & x_1z_1 & y_1z_1 \\ x_2^2 & y_2^2 & z_2^2 & x_2y_2 & x_2z_2 & y_2z_2 \\ x_3^2 & y_3^2 & z_3^2 & x_3y_3 & x_3z_3 & y_3z_3 \\ x_4^2 & y_4^2 & z_4^2 & x_4y_4 & x_4z_4 & y_4z_4 \\ x_5^2 & y_5^2 & z_5^2 & x_5y_5 & x_5z_5 & y_5z_5 \end{vmatrix} = 0.$$

For the rest, see [PenP86].

Next, we look at some problems dealing with fitting conics to given data. The following fact is used in justifying the constructions.

**3.6.1.4. Lemma.** If $\mathbf{C}_1$ and $\mathbf{C}_2$ are affine conics with equations $C_1(x,y) = 0$ and $C_2(x,y) = 0$, then

$$C_\lambda(x,y) = \lambda C_1(x,y) + (1-\lambda)C_2(x,y) = 0$$

or

$$C_\infty(x,y) = C_1(x,y) - C_2(x,y) = 0$$

is the equation of a conic $\mathbf{C}_\lambda$ that passes through the intersection points of the two given conics. If $\mathbf{C}_1$ and $\mathbf{C}_2$ have exactly four points of intersection, then the family $\mathbf{C}_\lambda$, $\lambda \in \mathbf{R}^*$, of conics consists of all the conics through these four points and each one is completely determined by specifying a fifth point on it.

**Proof.** See [PenP86].

Our design problems will also involve tangent lines and so we need to define those. Tangent lines play an important role when studying the geometry of curves. There are different ways to define them depending on whether one is looking at the curve from a topological or algebraic point of view. The definition we give here is specialized to conics. More general definitions will be encountered in Chapter 8 and 10. Our present definition is based on the fact that, at a point of a nondegenerate conic, the line that we would want to call the tangent line has the property that it is the only line through that point that meets the conic in only that point.

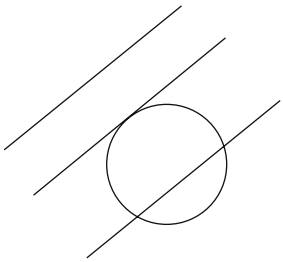


**Figure 3.22.** Possible line/circle intersections.

**3.6.1.5. Theorem.** Let **C** be a nondegenerate conic in $\mathbf{P}^2$. Any line **L** in $\mathbf{P}^2$ intersects **C** in 0, 1, or 2 points. Given any point **p** on **C**, there is one and only one line that intersects **C** in that single point.

**Proof.** This fact is easily checked if **C** is a circle. See Figure 3.22. The general case follows from the fact that any nonempty nondegenerate conic is projectively equivalent to a circle.

**Definition.** If a line **L** meets a nondegenerate conic **C** in a single point **p**, then **L** is called the *tangent line* to **C** at **p**. This definition applies to both the affine and projective conics.

**3.6.1.6. Theorem.** If a nondegenerate conic is defined in homogeneous coordinates by the equation

$$\mathbf{p}Q\mathbf{p}^T = 0,$$

then, in terms of homogeneous coordinates, the equation of the tangent line **L** to the conic at a point $\mathbf{P}_0$ is

$$\mathbf{p}Q\mathbf{p}_0{}^T = 0. \tag{3.68}$$

In particular, $[\mathbf{L}] = [\mathbf{p}_0 Q]$.

**Proof.** The line defined by equation (3.68) clearly contains $[\mathbf{p}_0]$. It therefore suffices to show that if another point satisfied equation (3.68), then the conic would be degenerate. See [PenP86].

**3.6.1.7. Corollary.** The equation of the tangent line at a point $(x_0,y_0)$ of a conic defined by equation (3.35) is

$$(ax_0 + hy_0 + f)x + (hx_0 + by_0 + g)y + fx_0 + gy_0 + c = 0.$$

**Proof.** Obvious.

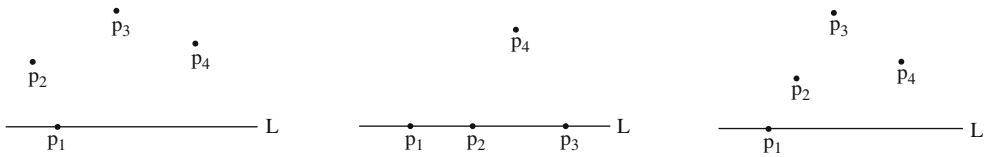



**Figure 3.23.** Conic design problem 1.

We now describe solutions to five conic design problems in the plane $\mathbf{R}^2$. The fact that the solutions are indeed correct follows easily from the above, in particular by repeated use of Lemma 3.6.1.4. See [PenP86] for details.

**Conic design problem 1:** To find the equation of the conic passing through four points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$ that has a given line $\mathbf{L}$ through one of these points as tangent line. Assume that at least two of the points do not lie on $\mathbf{L}$. If three points lie off $\mathbf{L}$, then no two of them are allowed to be collinear with the fourth. See Figure 3.23.

**Solution.** Assume that $\mathbf{L}$ is the tangent line at $\mathbf{p}_1$ and that $\mathbf{p}_2$ and $\mathbf{p}_3$ do not lie on $\mathbf{L}$. Let $\mathbf{L}_2$ be the line through $\mathbf{p}_1$ and $\mathbf{p}_2$, let $\mathbf{L}_3$ be the line through $\mathbf{p}_1$ and $\mathbf{p}_3$, and let $\mathbf{L}_4$ be the line through $\mathbf{p}_2$ and $\mathbf{p}_3$. Let $[\mathbf{L}] = [a,b,c]$ and $[\mathbf{L}_i] = [a_i,b_i,c_i]$. Define symmetric $3 \times 3$ matrices $Q_1$ and $Q_2$ by

$$Q_1 = \frac{1}{2}\Big((a,b,c)^T(a_4,b_4,c_4) + (a_4,b_4,c_t)^T(a,b,c)\Big) \quad \text{and}$$

$$Q_2 = \frac{1}{2}\Big((a_2,b_2,c_2)^T(a_3,b_3,c_3) + (a_3,b_3,c_3)^T(a_2,b_2,c_2)\Big).$$

Let $C_i(x,y) = 0$ be the quadratic equation associated to $Q_i$. Let $\mathbf{p}_4 = (x_4,y_4)$. If

$$C_\lambda(x,y) = \lambda C_1(x,y) + (1-\lambda)C_2(x,y),$$

then there is a unique $\lambda$ so that $C_\lambda(x_4,y_4) = 0$ and that is the equation of the conic we want.

**3.6.1.8. Example.** To find the conic that passes through the points $\mathbf{p}_1$, $\mathbf{p}_2 = (2,-2)$, $\mathbf{p}_3 = (2,2)$, $\mathbf{p}_4 = (5,0)$, and that has tangent line $\mathbf{L}$ at $\mathbf{p}_1$ for the case where $\mathbf{p}_1$ and $\mathbf{L}$ have the values

(a) $\mathbf{p}_1 = (-1,0)$, $\mathbf{L}: \ x+1=0$ (b) $\mathbf{p}_1 = (3,0)$, $\mathbf{L}: \ x-3=0$

**Solution for (a):** See Figure 3.24(a). We have the following equations for the lines $\mathbf{L}_i$:

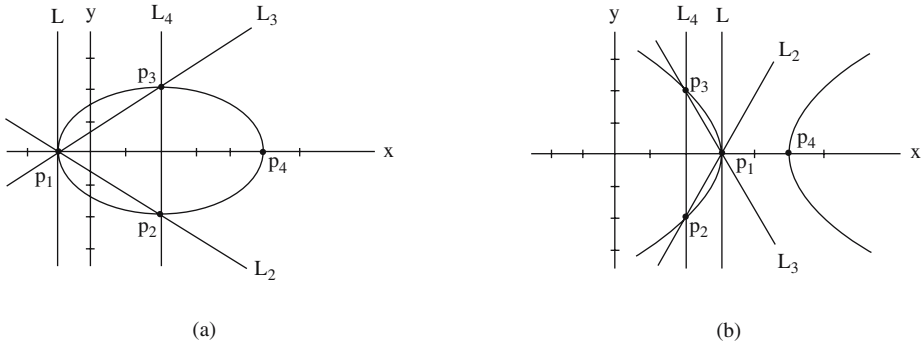



**Figure 3.24.** The conics that solve Example 3.6.1.6.

$$\mathbf{L}_2: \ 2x+3y+2=0$$
$$\mathbf{L}_3: \ 2x-3y+2=0$$
$$\mathbf{L}_4: \ x-2=0$$

Therefore,

$$C_\lambda(x,y)=\lambda(x+1)(x-2)+(1-\lambda)(2x+3y+2)(2x-3y+2).$$

Solving $C_\lambda(5,0) = 0$ for $\lambda$ gives $\lambda = 8/7$, so that our conic is the ellipse

$$C_{8/7}(x,y)=4(x-2)^2+9y^2-36=0.$$

**Solution for (b):** See Figure 3.24(b). We have the following equations for the lines $\mathbf{L}_i$:

$$\mathbf{L}_2: \ 2x+y-6=0$$
$$\mathbf{L}_3: \ 2x-y-6=0$$
$$\mathbf{L}_4: \ x-2=0$$

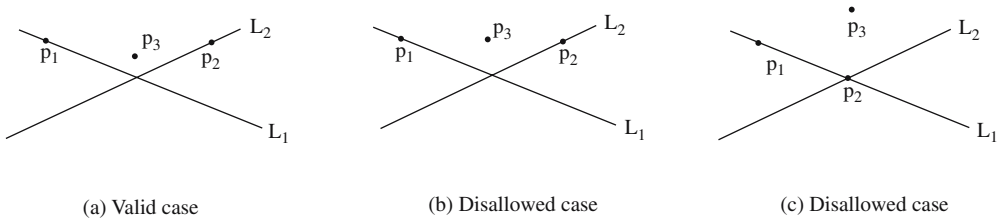
This time

$$C_\lambda(x,y)=\lambda(x-3)(x-2)+(1-\lambda)(2x+y-6)(2x-y-6).$$

Solving $C_\lambda(5,0) = 0$ for $\lambda$ gives $\lambda = 8/5$, so that our conic is the hyperbola

$$C_{8/5}(x,y)=-4(x-4)^2+3y^2+4=0.$$

**Conic design problem 2:** To find the equation of the conic passing through three points $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ that has two given lines $\mathbf{L}_1$ and $\mathbf{L}_2$ through two of these points as tangent lines. Assume that the three points are not collinear and that the intersec-




(a) Valid case (b) Disallowed case (c) Disallowed case

**Figure 3.25.** Conic design problem 2.

tion of the two lines is neither of the points where $\mathbf{L}_1$ and $\mathbf{L}_2$ are tangent to the conic. See Figure 3.25.

**Solution.** Assume that $\mathbf{L}_1$ and $\mathbf{L}_2$ are tangent lines at $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. Let $\mathbf{L}_3$ be the line through $\mathbf{p}_1$ and $\mathbf{p}_2$ and let $[\mathbf{L}_i] = [a_i,b_i,c_i]$. Define symmetric $3 \times 3$ matrices $Q_1$ and $Q_2$ by

$$Q_1 = \frac{1}{2}\left((a_1,b_1,c_1)^T(a_2,b_2,c_2) + (a_2,b_2,c_2)^T(a_1,b_1,c_1)\right) \quad \text{and} \quad Q_2 = (a_3,b_3,c_3)^T(a_3,b_3,c_3).$$

Let $C_i(x,y) = 0$ be the quadratic equation associated to $Q_i$. Let $\mathbf{p}_3 = (x_3,y_3)$. If

$$C_\lambda(x,y) = \lambda C_1(x,y) + (1-\lambda)C_2(x,y),$$

then there is a unique $\lambda$ so that $C_\lambda(x_3,y_3) = 0$ and that is the equation of the conic we want. Equivalently, if

$$Q_\lambda = \lambda Q_1 + (1-\lambda)Q_2,$$

then there is unique $\lambda$ so that $(x_3,y_3,1)Q_\lambda(x_3,y_3,1) = 0$.

**3.6.1.9. Example.** To find the conic that passes through the points $\mathbf{p}_1 = (4,-4)$ and $\mathbf{p}_2 = (4,4)$, has tangent lines

$$\mathbf{L}_1 : \ x - 2y + 4 = 0,$$
$$\mathbf{L}_2 : \ x + 2y + 4 = 0,$$

at those points, and also passes through the point

(a) $\mathbf{p}_3 = (0,0)$ (b) $\mathbf{p}_3 = (1,0)$ (c) $\mathbf{p}_3 = (-1,0)$

**Solution.** First note that the line $\mathbf{L}_3$ through $\mathbf{p}_1$ and $\mathbf{p}_2$ is clearly defined by

$$\mathbf{L}_3 : \ x - 4 = 0$$

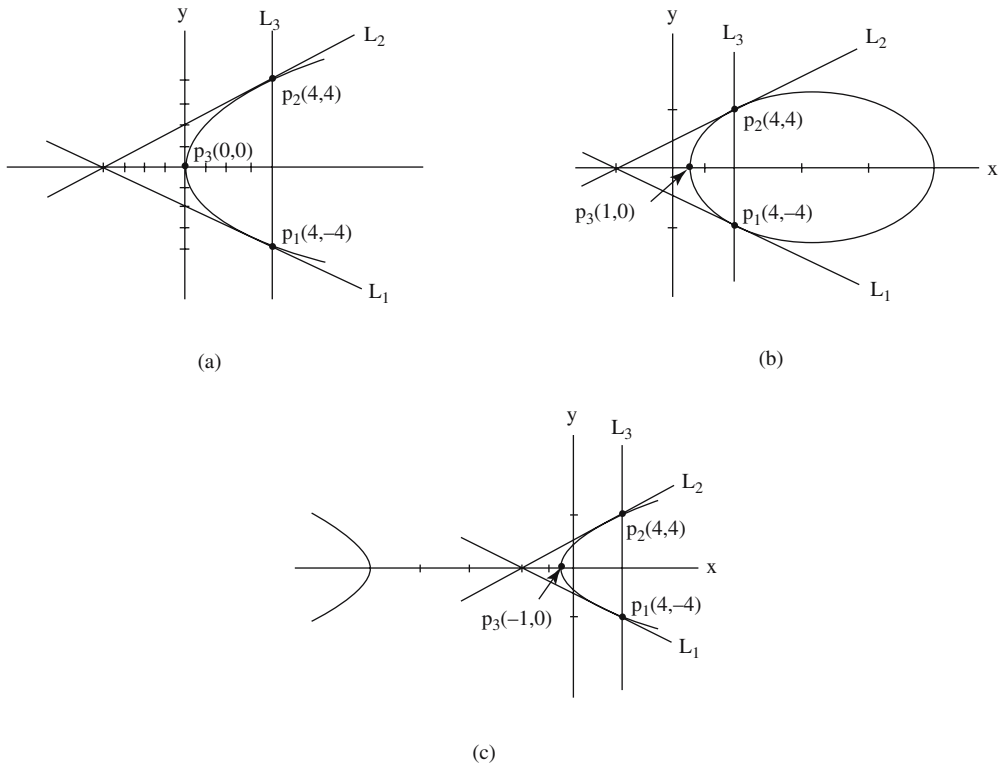



**Figure 3.26.** The conics that solve Example 3.6.1.7.

and

$$C_\lambda(x,y) = \lambda(x-2y+4)(x+2y+4) + (1-\lambda)(x-4)^2.$$

**Case** $\mathbf{p}_3 = (0,0)$: See Figure 3.26(a). The equation $C_\lambda(0,0) = 0$ leads to the impossible condition $l = 0$. This corresponds to the case $\lambda = \infty$. Therefore, the conic we are looking for is the parabola

$$C_\infty(x,y) = (x-2y+4)(x+2y+4) - (x-4)^2 = y^2 - 4x.$$

**Case** $\mathbf{p}_3 = (1,0)$: See Figure 3.26(b). The equation $C_\lambda(1,0) = 0$ leads to the solution $\lambda = -9/16$. This time our conic is the ellipse

$$C_{-9/16}(x,y) = 4x^2 - 68x + 9y^2 + 64 = 0.$$

**Case** $\mathbf{p}_3 = (-1,0)$: See Figure 3.26(c). The equation $C_\lambda(-1,0) = 0$ leads to the solution $\lambda = 25/16$. This time our conic is the hyperbola

$$C_{25/16}(x,y) = 4x^2 + 68x - 25y^2 + 64 = 0.$$

The next three design problems continue to decrease the number of specified points and increase the number of specified lines. Rather than solving them directly as we did for the first two problems, we shall note the duality of points and lines in projective space (as shown by equation (3.16) in Section 3.4) and essentially get our new solutions from those above using this duality. There will be little new that has to be proved. Mainly, we have to translate facts about points appropriately.

**Definition.** A *line conic* in the projective plane is a set of lines **L** that satisfy the equation

$$\mathbf{a}Q\mathbf{a}^T = 0,$$

where [**L**] = [**a**] and Q is a symmetric 3 × 3 matrix.

**3.6.1.10. Theorem.** The set of tangent lines **L** to the nondegenerate point conic defined by

$$\mathbf{p}Q\mathbf{p}^T = 0$$

is the line conic defined by

$$\mathbf{a}Q^{-1}\mathbf{a}^T = 0,$$

where [**L**] = [**a**].

**Proof.** This theorem is an easy consequence of Theorem 3.6.1.5. We have replaced the point [**p**] by the line [**a**], where **a** = **p**Q, which is tangent to the conic at **p**. Turning this around, the point [$\mathbf{a}Q^{-1}$] corresponds to the line [**a**].

**Conic design problem 3:** To find the equation of the conic that passes through two points and is tangent to three lines, such that one of the lines passes through the first point, another passes through the second point, and the third line is arbitrary. The lines are not allowed to be concurrent and the neither of the first two lines can contain both points. See Figure 3.27(a).

**Solution.** We dualize the solution to design problem 2. Assume that the conic passes through $\mathbf{p}_1$ and $\mathbf{p}_2$ and has tangent lines $\mathbf{L}_1$ and $\mathbf{L}_2$ at those points. Let $\mathbf{L}_3$ be the other tangent line. Let $\mathbf{p}_3$ be the intersection of $\mathbf{L}_1$ and $\mathbf{L}_2$, and let $\mathbf{p}_i$ have homogeneous coordinates $[x_i,y_i,z_i]$. Define symmetric 3 × 3 matrices $Q_1$ and $Q_2$ by

$$Q_1 = \frac{1}{2}\left((x_1,y_1,z_1)^T(x_2,y_2,z_2) + (x_2,y_2,z_2)^T(x_1,y_1,z_1)\right) \quad \text{and} \quad Q_2 = (x_3,y_3,z_3)^T(x_3,y_3,z_3).$$

Let $[\mathbf{L}_3] = [a_3,b_3,c_3]$. If

$$Q_\lambda = \lambda Q_1 + (1-\lambda)Q_2,$$

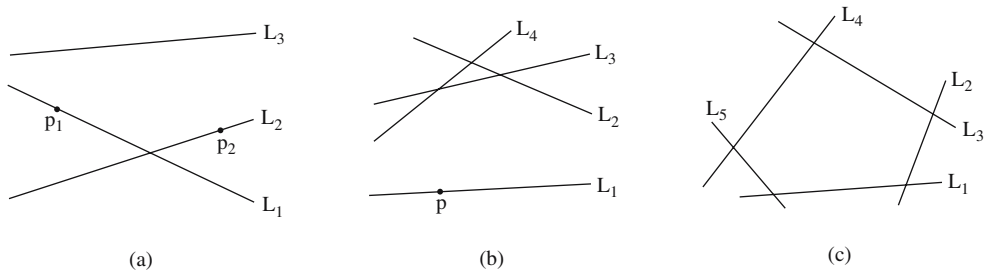



**Figure 3.27.** Conic design problems 3, 4, and 5.

then there is a unique $\lambda$ so that $(a_3,b_3,c_3)Q_\lambda(a_3,b_3,c_3)^T = 0$ and $Q_\lambda^{-1}$ is the symmetric matrix that defines the (point) conic we want.

We leave the solutions of the next two design problems as exercises for the reader (Exercise 3.6.1.6 and 3.6.1.7). Note however that the requirements are stronger than for the point analog because we need $Q^{-1}$. See [PenP86].

**Conic design problem 4:** To find the equation of the conic that passes through one point, is tangent to a line through this point, and is also tangent to three other lines. None of the last three lines may contain the given point and no three of the four lines are allowed to be concurrent. See Figure 3.27(b).

**Conic design problem 5:** To find the equation of the conic that is tangent to five lines, no three of which are allowed to be concurrent. See Figure 3.27(c).

## 3.7 Quadric Surfaces

This section defines quadric surfaces and discusses the equations that define them.

**Definition.** A *quadratic surface* is a set **S** of points in $\mathbf{R}^n$ whose coordinates satisfy an equation

$$q(x_1, x_2, \ldots, x_n) = 0,$$

where q is a quadratic polynomial function in n variables and not all the coefficients of the monomial terms in q of total degree 2 vanish. A quadratic surface in $\mathbf{R}^3$ is called a *quadric surface*.

By definition, quadric surfaces are the solution sets of arbitrary quadratic equation in three variables of the form

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0. \tag{3.69}$$

where at least one of the six coefficients a, b, c, d, e, or f is nonzero. Such equations define surfaces in general.

In trying to analyze the type of surface that equation (3.69) can give rise to, the hard part is getting rid of the xy, xz, and yz cross terms. One can, like in the 2-variable case, use a rotation to change the coordinate system to one in which the equation has the form

$$a'x'^2 + b'y'^2 + c'z'^2 + g'x' + h'y' + i'z' + j' = 0. \tag{3.70}$$

One could, for example, use the roll-pitch-yaw representation for a rotation about the origin. This involves three unknowns. Substituting the rotated points into (3.69) and then setting the coefficients of the cross terms to 0, would give three equations in three unknowns which could be solved, but this is starting to get too complicated and messy.

The more elegant way to eliminate the cross terms in equation (3.69) or quadratic equation in any number of variables is to use the theory of quadratic forms. Like in the 2-variable case let

$$q(x,y,z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz$$

be the associated quadratic form. It follows again from Theorem 1.9.10 that q is diagonalizable and there is a suitable coordinate system in which equation (3.69) has the form (3.70). If any quadratic term is present in (3.70), then the corresponding linear term, if there is one, can be eliminated by completing the square similar to the way it was done in Section 3.6. The analysis depends on whether a′, b′, or c′ are zero, and if not, on whether they are positive or negative. The resulting cases are easy to analyze and lead to the following theorem.

**3.7.1. Theorem.** Any equation of the form (3.69) can be transformed via a rigid motion into one of the following fourteen types of equations (equivalently, there is a coordinate system, called the *natural coordinate system* for the quadric, with respect to which equation (3.69) has the following form):

(1) Ellipsoid: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$

(2) Hyperboloid of one sheet: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$

(3) Hyperboloid of two sheets: $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$

(4) Empty set: $-\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$

(5) Point: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 0$

(6) Cone: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 0$

(7) Elliptic or hyperbolic paraboloid: $\dfrac{x^2}{a^2} \pm \dfrac{y^2}{b^2} = cz$, where $c \neq 0$

(8) Elliptic or hyperbolic cylinder: $\frac{x^2}{a^2} \pm \frac{y^2}{b^2} = 1$

(9) Empty set: $-\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

(10) Line: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$

(11) Two intersecting planes: $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$

(12) Parabolic cylinder: $\frac{x^2}{a^2} = by$, where $b \neq 0$

(13) One or two parallel planes: $x^2 = a^2$

(14) Empty set: $x^2 = -a^2$, where $a \neq 0$

**Proof.** See [Divi75] for more details.

Figure 3.28 shows a few of the types of solutions to equation (3.69). Most of the solutions are nice quadratic surfaces, but some solutions are degenerate.

**Definition.** The coordinate planes of a quadric in its natural coordinate system that are planes of symmetry for the quadric are called the *principal planes* of the quadric.

Ellipsoids and hyperboloids of one or two sheets have three principal planes. The elliptic paraboloid has two. Below are some geometric properties of quadrics. See [HilC99].

(1) A quadric and a plane always intersect in a conic curve.

(2) If $a = b$ for the surfaces (1)–(3), (6), the elliptic paraboloid in (7), and the elliptic cylinder in (8) of Theorem 3.7.1, then we get surfaces of revolution. For example, the ellipsoid is then just an ellipse revolved about its major or minor axis. The hyperboloid of two sheets is obtained by revolving a hyperbola about its axis. The hyperboloid of one sheet is obtained by revolving a hyperbola about the perpendicular bisector of the segment between its two foci. This type of surface can also be obtained by rotating a line skew with the axis of revolution about that axis. The general case of the just mentioned surfaces where $a \neq b$ is gotten by starting with a surface of revolution of that type, fixing a plane through its axis, and pulling all the points of the surface away from that plane in such a way that the distances of the points from the plane change by a fixed ratio.

(3) The elliptic, hyperbolic, and parabolic cylinders, the hyperboloid of one sheet, and the hyperbolic paraboloid are ruled surfaces in that they are swept out by a family of straight lines. (See Chapter 9 for more on ruled surfaces.) The last two are doubly ruled surfaces, that is, they are swept out by two distinct families of straight lines.

(4) There is a string construction for the ellipsoid. We use an ellipse and hyperbola in orthogonal planes, where the foci $\mathbf{F}_1$ and $\mathbf{F}_2$ of the ellipse are the vertices of the hyperbola and the foci of the hyperbola are the vertices $\mathbf{V}_1$ and $\mathbf{V}_2$ of the ellipse. See Figure 3.29. Tie a string to one of the vertices of the ellipse, say $\mathbf{V}_1$. Now loop the

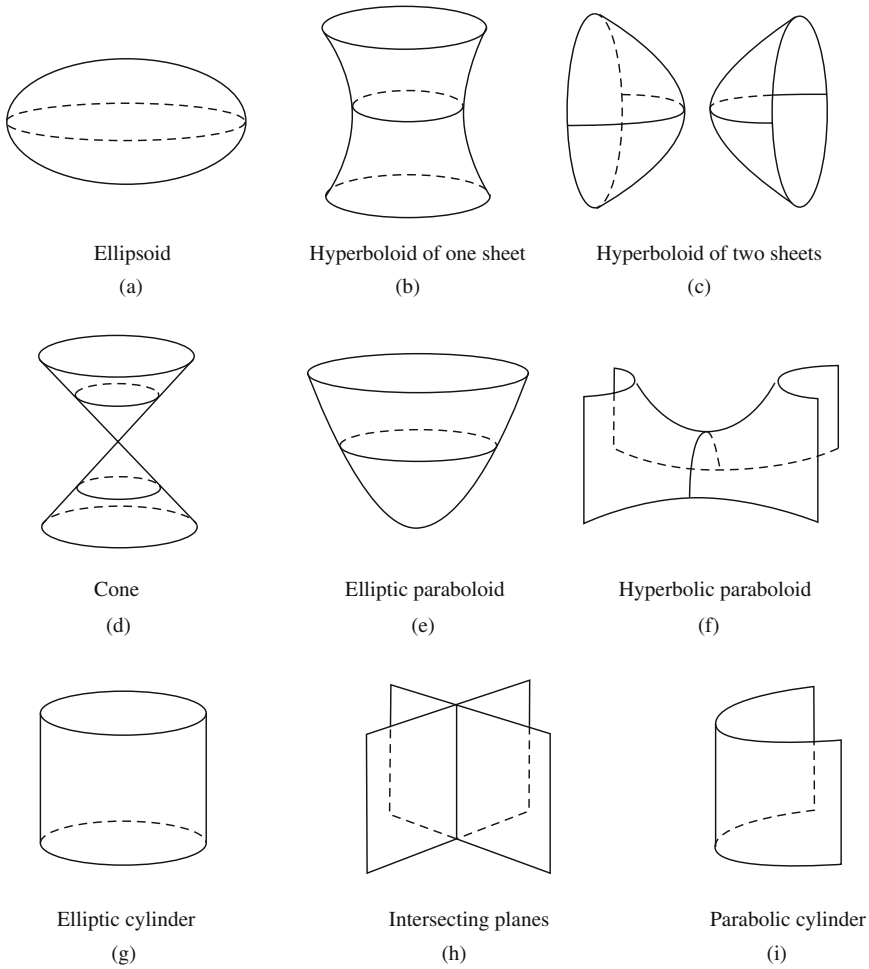



**Figure 3.28.** Some quadric surfaces.

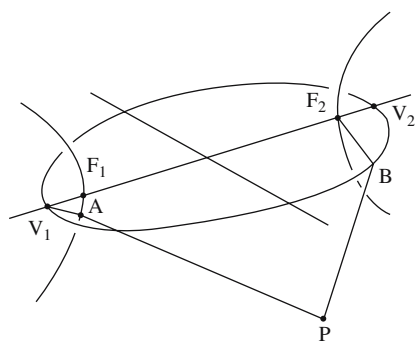



**Figure 3.29.** A string construction for the ellipsoid.

string around the near branch of the hyperbola with vertex $\mathbf{F}_1$ (at point **A** in the figure), then around the outside of the ellipse (at point **B** in the figure), and finally attach it to $\mathbf{F}_2$. If we pull the string tight at some point **P**, then the locus of those points **P** will trace out one quarter of an ellipsoid. By attaching the string at corresponding different locations we can get the rest of the ellipsoid. Because of the similarity of this construction with that for the ellipse, one makes the following definitions.

**Definition.** The ellipse and hyperbola used to construct the ellipsoid are called the *focal curves* (the *focal ellipse* and *focal hyperbola*) of the ellipsoid. In general, given any quadric, we say that two conics in orthogonal principal planes for this quadric are *focal curves* for the quadric if they are confocal with the intersection of the principal planes with the quadric. Two quadric surfaces with the same focal curves are called *confocal*.

Only ellipsoids and hyperboloids of one or two sheets have focal curves. The family of all confocal quadrics of one of those three types that have a fixed pair of focal curves fill up all of space. The tangent planes of the three just-mentioned confocal families are mutually orthogonal at a point of intersection. (For a precise definition of a tangent plane see Chapter 8.)

Next, we state the analog of Theorem 3.6.4 for the surface case. We rewrite equation (3.69) as

$$ax^2 + by^2 + cz^2 + 2hxy + 2fyz + 2gzx + 2lx + 2my + 2nz + d = 0. \tag{3.71}$$

Define matrices A and B by

$$A = \begin{pmatrix} a & h & f & l \\ h & b & g & m \\ f & g & c & n \\ l & m & n & d \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} a & h & f \\ h & b & g \\ f & g & c \end{pmatrix}. \tag{3.72}$$

**3.7.2. Theorem.** Define $\Delta$, D, I, and J for equation (3.71) by

$$\Delta = \det(A) = \begin{vmatrix} a & h & f & l \\ h & b & g & m \\ f & g & c & n \\ l & m & n & f \end{vmatrix}, \quad D = \det(B) = \begin{vmatrix} a & h & f \\ h & b & g \\ f & g & c \end{vmatrix},$$

$$I = \operatorname{tr}(B) = a + b + c, \quad \text{and} \quad J = bc + ca + ab - f^2 - g^2 - h^2.$$

(1) The quantities $\Delta$, D, I, and J are invariant under a change of coordinates via rigid motions (translations and/or rotations).
(2) If $D \neq 0$, then let $r_1$, $r_2$, and $r_3$ be the nonzero eigenvalues of the matrix B in (3.72).

(a) $\Delta < 0$:

I D > 0, J > 0: We have an ellipsoid.
Not both I D > 0 and J > 0: We have an hyperboloid of two sheets.

(b) $\Delta > 0$:

I D > 0, J > 0: The empty set.
Not both I D > 0 and J > 0: We have an hyperboloid of one sheet.

(c) $\Delta = 0$:

I D > 0, J > 0: A single point.
Not both I J > 0 and J > 0: We have a cone.

In case (a) and (b), equation (3.71) can be reduced to

$$r_1x^2 + r_2y^2 + r_3z^2 + \frac{\Delta}{D} = 0.$$

and in case (c) to

$$r_1x^2 + r_2y^2 + r_3z^2 = 0.$$

(3) If D = 0, then equation (3.71) defines a paraboloid or cylindrical surface generated by a conic in a plane (types (8) and (12) in Theorem 3.7.1) unless the surface is degenerate. More precisely, let $r_i$ be the nonzero eigenvalues of the matrix B in (3.72).

(a) $\Delta \neq 0$:

Equation (3.71) can be reduced to

$$r_1x^2 + r_2y^2 + 2\sqrt{\frac{-\Delta}{r_1r_2}}z = 0.$$

$\Delta < 0$, J > 0: We have an elliptic paraboloid.
$\Delta > 0$, J < 0: We have an hyperbolic paraboloid.

(b) $\Delta = 0$:

If

$$\begin{vmatrix} a & h & l \\ h & b & m \\ l & m & d \end{vmatrix} = \begin{vmatrix} a & f & l \\ f & c & n \\ l & n & d \end{vmatrix} = \begin{vmatrix} b & g & m \\ g & c & n \\ m & n & d \end{vmatrix} = 0,$$

then the surface is degenerate and reduces to a plane or a pair of planes. Otherwise, it is a cylindrical surface of one of the following types:

J > 0: a real elliptic cylinder or the empty set.
J < 0: a hyperbolic cylinder.
J = 0: a parabolic cylinder.

Alternatively, let k be the number of nonzero eigenvalues of the matrix B in (3.72). Then equation (3.71) can be reduced to one of the following:

$k=1$: $r_1x^2+2my=0$ (parabolic cylinder) , or
$r_1x^2+d=0$ (two planes)

$k=2$: $r_1x^2+r_2y^2+d=0$
(If solutions exist, then we get an elliptic or hyperbolic cylinder if $d\neq 0$ and two planes if $d=0$.)

**Proof.** See [Eise39].

**3.7.3. Theorem.** The equation of the tangent plane to a quadric surface at a point $(x_0,y_0,z_0)$ defined by equation (3.71) is

$$(ax_0+hy_0+fz_0+l)x+(by_0+hx_0+gz_0+m)y+(cz_0+fx_0+gy_0+n)z$$
$$+lx_0+my_0+nz_0+d=0.$$

**Proof.** See [Eise39]. (We again refer the reader to Chapter 8 for a precise definition of a tangent plane).

For a classification of quadratic surfaces in $\mathbf{R}^n$ see [PetR98].

## 3.8 Generalized Central Projections

The standard central projections as defined in Section 3.2 have a center that is a point. When dealing with higher-dimensional spaces it is sometimes convenient to allow the "center" to be an arbitrary plane.

**Definition.** Let $\mathbf{O}^{n-k-1}$ be a fixed $(n-k-1)$-dimensional plane in $\mathbf{R}^n$. If $\mathbf{Y}^k$ is a k-dimensional plane in $\mathbf{R}^n$, define a map

$$\pi_{\mathbf{O}}:\ \mathbf{R}^n \to \mathbf{Y}$$

by

$$\pi_{\mathbf{O}}(\mathbf{p}) = \mathrm{aff}(\mathbf{O},\mathbf{p}) \cap \mathbf{Y},\ \text{if aff}(\mathbf{O},\mathbf{p}) \text{ intersects } \mathbf{Y} \text{ in a single point,}$$
$$= \text{undefined, otherwise.}$$

The map $\pi_{\mathbf{O}}$ is called the *generalized central projection with center* $\boldsymbol{O}$ *of* $\boldsymbol{R}^n$ *to the plane* $\boldsymbol{Y}$. If $\mathbf{X}^k$ is another k-dimensional plane in $\mathbf{R}^n$, then the restriction of $\pi_{\mathbf{O}}$ to $\mathbf{X}$, $\pi_{\mathbf{O}}|\mathbf{X}$ : $\mathbf{X} \to \mathbf{Y}$, is called the *generalized perspective transformation* or *generalized perspectivity from* $\boldsymbol{X}$ *to* $\boldsymbol{Y}$ *with center* $\boldsymbol{O}$.

**3.8.1. Theorem.** The generalized perspectivity $\pi_{\mathbf{O}}|\mathbf{X}$ from $\mathbf{X}$ to $\mathbf{Y}$ with center $\mathbf{O}$ is a projective transformation.

**Proof.** It is not hard to show that $\pi_{\mathbf{O}}|\mathbf{X}$ is a composition of ordinary central projections and parallel projections.

Our next task is to show that working with generalized central projections is just as easy as with ordinary central projections. We extend the notion of a frame and borrow some of the next terminology from computer graphics.

**Definition.** A *generalized frame* in $\mathbf{R}^n$ is a tuple $F = (\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{k+1},\mathbf{p})$, where the $\mathbf{u}_i$ form an orthonormal set of vectors in $\mathbf{R}^n$ and $\mathbf{p}$ is point of $\mathbf{R}^n$. The $(n - k - 1)$-dimensional plane $\mathbf{B}$ for which F is a point-normals representation is called the *base plane* of F and denoted by base(F). If $\mathbf{u}_{k+2}$, $\mathbf{u}_{k+3}$, . . . , and $\mathbf{u}_n$ are an orthonormal basis for $\mathbf{B}$, then the frame $G = (\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n,\mathbf{p})$ is called an *augmented frame* for F. The vector $\mathbf{u}_{k+1}$ is called the *view direction* of F. Any k-dimensional plane $\mathbf{V}$ that passes through a point $\mathbf{o} = \mathbf{p} + d\mathbf{u}_{k+1}$, for some $d > 0$, and has basis $\mathbf{u}_1$, $\mathbf{u}_2$, . . . , and $\mathbf{u}_k$ is called a *view plane* for F. The view plane $\mathbf{V}$ is said to be a *distance d in front of* $\mathbf{B}$. The frame $(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k,\mathbf{o})$ is called the *view plane coordinate system* and the point $\mathbf{o}$ is called the *origin* of the view plane.

Before moving on to the general case, it is helpful to work out the computational details in the special case of $\mathbf{R}^3$. Assume that $\mathbf{p} = (0,-1/a,0)$, $F = (\mathbf{e}_1,\mathbf{e}_2,\mathbf{p})$, and that the view plane $\mathbf{V}$ is the x-axis and has origin $\mathbf{0}$. See Figure 3.30. By an argument similar to the one in Section 3.5.1 one can easily show that the generalized central projection with respect to F and $\mathbf{V}$

$$C_a : \mathbf{R}^3 \to \mathbf{V}$$

is defined by

$$C_a(x,y,z) = (x',0,0) = \left(\frac{x}{ay+1},0,0\right).$$

The map $C_a$ can also be described as an orthogonal projection parallel to the z-axis to the x-y plane followed by an ordinary central projection of $\mathbf{R}^2$ onto the x-axis from

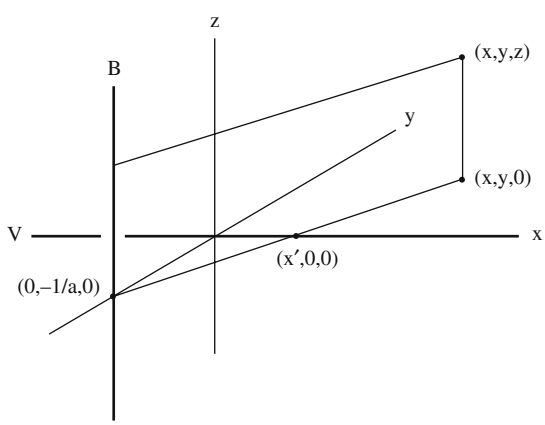



**Figure 3.30.** A basic generalized central projection in $\mathbf{R}^3$.

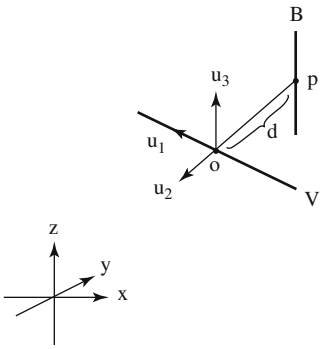



**Figure 3.31.** A generalized central projection in $\mathbf{R}^3$.

the point (0,–1/a,0). In other words, using homogeneous coordinates the map $C_a$ is defined by the matrix

$$M_a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Now for an arbitrary generalized frame $F = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{p})$ in $\mathbf{R}^3$, let $\mathbf{V}$ be a view plane for F that is a distance d in front of $\mathbf{B}$ = base(F) and that has origin $\mathbf{o}$. See Figure 3.31. Let $\pi$ be the generalized central projection of $\mathbf{R}^3$ on $\mathbf{V}$ with center $\mathbf{p}$. Suppose that we would like to determine a formula for $\pi(\mathbf{p})$ **in view plane coordinates** for an arbitrary point $\mathbf{p} \in \mathbf{R}^3$. Choose an orthonormal basis $\mathbf{u}_3$ for $\mathbf{B}$ and let $G = (\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{p})$ be the corresponding augmented frame for F. The map $G^{-1}$ maps world coordinates into the coordinates of the frame G. Let T be the translation $\mathbf{q} \to \mathbf{q} - (0,d,0)$. Then $\pi = C_{1/d}TG^{-1}$. If we use homogeneous coordinates, the $\pi$ can be expressed in terms of a matrix, like in Section 3.5.1. Actually, because of all the zeros in the matrix $M_a$, the computations can be simplified. To compute $\pi(\mathbf{q})$ we only need to compute the dot product of $\mathbf{q}$ with $\mathbf{u}_1$ and keep track of that. More precisely,

$$\pi(\mathbf{q}) = \left(\frac{d(\mathbf{q}-\mathbf{p})\bullet\mathbf{u}_1}{(\mathbf{q}-\mathbf{p})\bullet\mathbf{u}_2},0,0\right).$$

Now back to the general problem in $\mathbf{R}^n$. Given a generalized frame $F = (\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_{k+1},\mathbf{p})$ in $\mathbf{R}^n$, the problem that interests us is to determine the generalized central projection $\pi$ with center $\mathbf{B}$ = base(F) of $\mathbf{R}^n$ to a view plane $\mathbf{V}$ that is a distance d in front of $\mathbf{B}$. There is again the important special case where $\mathbf{p} = (-1/a)\mathbf{e}_{k+1}$, $F = (\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_{k+1},\mathbf{p})$, and the view plane $\mathbf{V}$ is $\mathbf{R}^k$ with origin $\mathbf{0}$. Using homogeneous coordinates the generalized central projection $C_a$ in this case is defined by the matrix

$$M_a = \begin{pmatrix} \overbrace{\begin{matrix} 1 & 0 \\ \ddots & \\ 0 & 1 \end{matrix}}^{k} & & & \\ & 0 & & 0 & 0 \\ & & \overbrace{\begin{matrix} 0 & 0 \\ \ddots & \\ 0 & 0 \end{matrix}}^{n-k-1} & & a \\ 0 & & & 0 & \\ & & & & 1 \end{pmatrix}$$

The general case $\pi$ is reduced to this special case using a motion like before. Choose an orthonormal basis for **B** and let $G = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \mathbf{p})$ be the corresponding augmented frame for F. The map $G^{-1}$ maps world coordinates into the coordinates of the frame G. Let T be the translation $\mathbf{q} \to \mathbf{q} - d\mathbf{e}_{k+1}$. Then $\pi = C_{1/d}TG^{-1}$. The affine version of the map is

$$\pi(\mathbf{q}) = \left( d\frac{(\mathbf{q}-\mathbf{p})\bullet \mathbf{u}_1}{(\mathbf{q}-\mathbf{p})\bullet \mathbf{u}_{k+1}}, \ldots, d\frac{(\mathbf{q}-\mathbf{p})\bullet \mathbf{u}_k}{(\mathbf{p}-\mathbf{p})\bullet \mathbf{u}_{k+1}}, 0, \ldots, 0 \right).$$

## 3.9 The Theorems of Pascal and Brianchon

It did not seem appropriate to leave the subject of projective geometry without mentioning two well-known and beautiful theorems.

**3.9.1. Theorem.**

(1) (Pascal's Theorem) If the vertices **A**, **B**, **C**, **D**, **E**, and **F** of a hexagon in $\mathbf{P}^2$, no three of which are collinear, lie on a nondegenerate conic, then the pairs of opposite sides intersect in collinear points, that is, the intersection points

$$\mathbf{X} = \mathbf{L_{AB}} \cap \mathbf{L_{DE}}, \quad \mathbf{Y} = \mathbf{L_{BC}} \cap \mathbf{L_{FE}}, \quad \text{and} \quad \mathbf{Z} = \mathbf{L_{AF}} \cap \mathbf{L_{CD}}$$

are collinear, where $\mathbf{L_{PQ}}$ is the line through points **P** and **Q**. See Figure 3.32.

(2) Conversely, if the pairs of opposite sides of a hexagon in $\mathbf{P}^2$, no three of whose vertices are collinear, intersect in collinear points, then the vertices of the hexagon lie on a nondegenerate conic.

**Proof.** See [Gans69] or [Fari95].

Pascal proved his theorem in 1640. There are affine versions of these theorems but they are not as elegant because one has to add assumptions that intersections

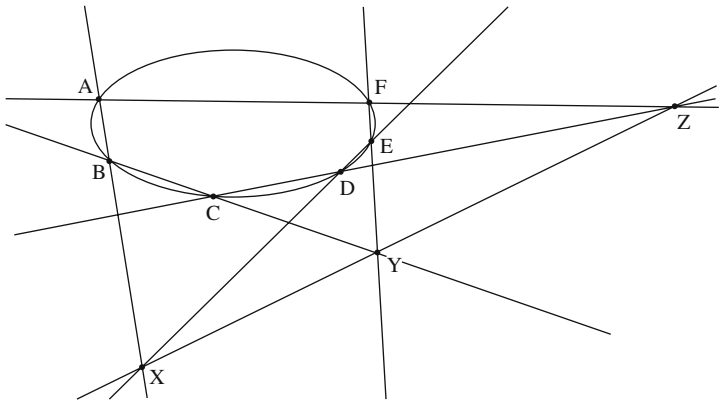



**Figure 3.32.** Pascal's theorem.

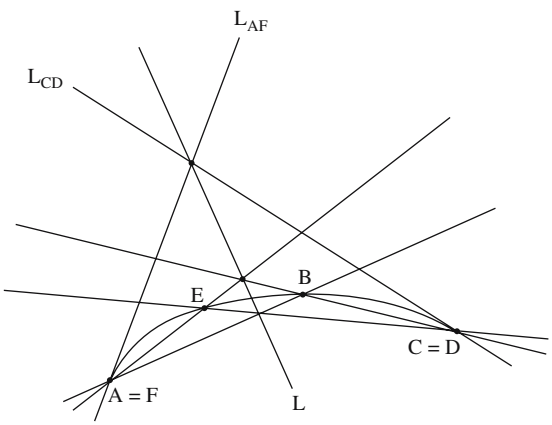



**Figure 3.33.** Points on a conic given three points and two tangents.

exist. The book [HilC99] has a nice discussion about the connection between Pascal's theorem and other results.

The converse of Pascal's theorem is actually more interesting because it allows one to construct any number of points on a conic through five points no three of which are collinear. Also, by letting the points **A** and **F** and the points **C** and **D** coalesce we get a construction for the points on a conic given three points and the tangent lines at the first two points. For example, in Figure 3.33 (which uses notation compatible with that in Theorem 3.9.1) we are given the three points $\mathbf{A} = \mathbf{F}$, $\mathbf{C} = \mathbf{D}$, and **B** and tangent lines $\mathbf{L_{AF}}$ and $\mathbf{L_{CD}}$ at the points **A** and **C**, respectively. The figure shows how an arbitrary line **L** through the intersection of $\mathbf{L_{AF}}$ and $\mathbf{L_{CD}}$ determines a unique new point **E** on the conic. Note that since the line **L** can be parameterized by the angle that it makes with the line $\mathbf{L_{AF}}$, our construction also produces a parameterization of the points on the conic.

The dual of Pascal's Theorem is called Brianchon's Theorem after C.J. Brianchon, who proved it in 1806 before the principle of duality in $\mathbf{P}^2$ was formulated. Of course,

given that principle, there would have been nothing to prove. The dual of Theorem 3.9.1(2) holds also, again by the principle of duality.

## 3.10 The Stereographic Projection

The subject matter of this last section is not really part of projective geometry as such but of geometry in a general sense. The map we shall describe shows up in many places, including topology and complex analysis. It has many interesting properties but we shall only take time to discuss those that are relevant to this book. A good reference for more information is [HilC99].

**Definition.** The *stereographic projection* of the n-sphere

$$p_n : \mathbf{S}^n - \{\mathbf{e}_{n+1}\} \to \mathbf{R}^n$$

is defined by

$$p_n(x_1, x_2, \ldots, x_{n+1}) = \left(\frac{x_1}{1 - x_{n+1}}, \frac{x_2}{1 - x_{n+1}}, \ldots, \frac{x_n}{1 - x_{n+1}}\right).$$

Note that $p_n|\mathbf{S}^{n-1}$ is the identity map on $\mathbf{S}^{n-1}$. The map $p_n$ can be described geometrically as follows: If $\mathbf{x} \in \mathbf{S}^n - \{\mathbf{e}_{n+1}\}$ and if $\mathbf{L_x}$ is the ray that starts at $\mathbf{e}_{n+1}$ and passes through $\mathbf{x}$, then $p_n(\mathbf{x})$ is the point that is the intersection of $\mathbf{L_x}$ with $\mathbf{R}^n$. See Figure 3.34 for the case n = 2. It is easy to check that $p_n$ is one-to-one and onto, so that we may think of $\mathbf{S}^n$ as

$$\mathbf{R}^n_\infty = \mathbf{R}^n \cup \{\infty\},$$

where $\infty$ is the "point of $\mathbf{R}^n$ at infinity." Using terminology introduced later in Chapter 5, $\mathbf{S}^n$ can be thought of as the one-point compactification of $\mathbf{R}^n$. The map $p_n$ extends to a one-to-one and onto map

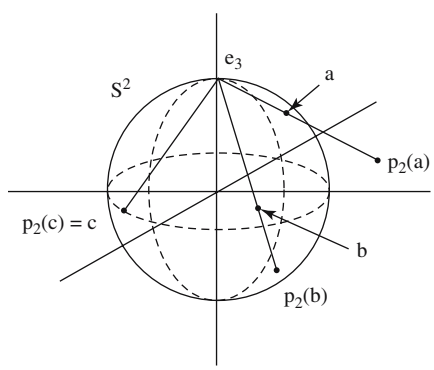



**Figure 3.34.** The stereographic projection.

$$p_n : \mathbf{S}^n \to \mathbf{R}^n_\infty$$

by mapping $\mathbf{e}_{n+1}$ to $\infty$. This map is also called the *stereographic projection* of the n-sphere.

The identification of the two space $\mathbf{S}^n$ and $\mathbf{R}^n_\infty$ using $p_n$ gives us a one-to-one correspondence between maps of $\mathbf{S}^n$ and $\mathbf{R}^n_\infty$. Specifically, for any map

$$h : \mathbf{S}^n \to \mathbf{S}^n$$

define

$$p_n(h) : \mathbf{R}^n_\infty \to \mathbf{R}^n_\infty$$

by

$$p_n(h) = p_n \circ h \circ p_n^{-1}.$$

Alternatively, $p_n(h)$ is the unique map that makes the following diagram commutative:

$$\begin{array}{ccc} S^n & \xrightarrow{h} & S^n \\ p_n \downarrow & & \downarrow p_n. \\ \mathbf{R}^n_\infty & \xrightarrow[p_n(h)]{} & \mathbf{R}^n_\infty \end{array}$$

**3.10.1. Theorem.**

(1) If $\mathbf{X}$ is an k-dimensional sphere in $\mathbf{S}^n$ that misses the point $\mathbf{e}_{n+1}$, then $\mathbf{X}' = p_n(\mathbf{X})$ is a k-dimensional sphere in $\mathbf{R}^n$.
(2) If $\mathbf{X}$ is an k-dimensional sphere in $\mathbf{S}^n$ through the point $\mathbf{e}_{n+1}$, then $\mathbf{X}' = p_n(\mathbf{X})$ is a k-dimensional plane in $\mathbf{R}^n$.

**Proof.** Consider the case of circles and n = 2. The argument for part (1) proceeds as follows. Let $\mathbf{X}$ be a circle in $\mathbf{S}^2$ that does not pass through $\mathbf{e}_3$. Figure 3.35 shows a vertical slice of the three-dimensional picture. The points $\mathbf{A}$ and $\mathbf{B}$ are points of $\mathbf{X}$ and $\mathbf{A}'$ and $\mathbf{B}'$ are the image of $\mathbf{A}$ and $\mathbf{B}$, respectively, under the stereographic projection. The tangent planes at the points of $\mathbf{X}$ envelop a cone with vertex $\mathbf{C}$. One can show that the image $\mathbf{C}'$ of $\mathbf{C}$ under the stereographic projection is then the center of the circle $\mathbf{X}'$. See [HilC99]. Part (2) follows from the fact that a circle through $\mathbf{e}_3$ is the intersection of a plane with the sphere. The general case of spheres and arbitrary k-dimensional spheres is proved in a similar fashion.

If we consider a plane as a "sphere through infinity," then Theorem 3.10.1 can be interpreted as saying that the stereographic projection takes spheres to spheres. With this terminology we can now also talk about *sphere-preserving transformations* of both $\mathbf{S}^n$ and $\mathbf{R}^n_\infty$. (Note that in $\mathbf{R}^n$ these would be the maps that send a sphere to a sphere **or** a plane and a plane to a plane **or** a sphere. We would have had a problem talking about such "sphere-preserving" transformations in $\mathbf{R}^n$ because we would have to allow

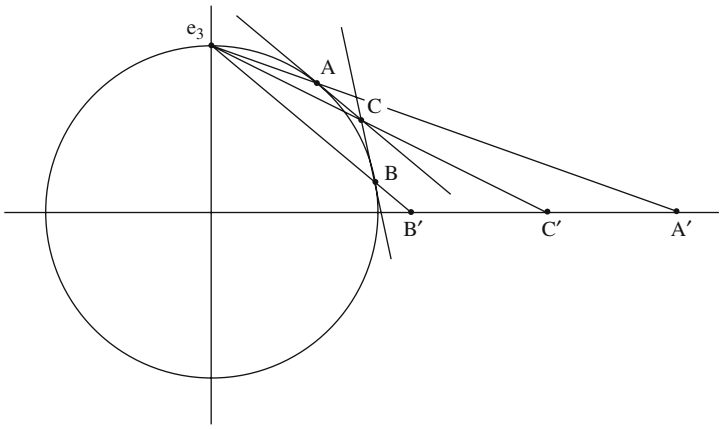



**Figure 3.35.** The stereographic projection maps circles to circles.

these transformations to not be defined at a point and not onto a point. We had a similar problem with affine projective transformations.) The sphere-preserving transformations of $\mathbf{R}^n_\infty$ are easy to characterize. First, any similarity of $\mathbf{R}^n$ extends to a map of $\mathbf{R}^n_\infty$ to itself by sending $\infty$ to $\infty$. Call such a map of $\mathbf{R}^n_\infty$ an *extended similarity*.

**3.10.2. Theorem.**

(1) An extended similarity of $\mathbf{R}^n_\infty$ is a sphere-preserving map. Conversely, every sphere-preserving maps of $\mathbf{R}^n_\infty$ that leave $\infty$ fixed is an extended similarity.
(2) An arbitrary sphere-preserving maps of $\mathbf{R}^n_\infty$ is a composition of an extended similarity and/or a map $p_n(h)$, where h is a rotation of $\mathbf{S}^n$ around a great circle through $\mathbf{e}_{n+1}$.

**Proof.** See [HilC99].

Another interesting and important property of the stereographic projection is that there is a sense in which it preserves angles. Let $\mathbf{p}$ be any point of $\mathbf{S}^n$ other than $\mathbf{e}_{n+1}$. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent tangent vector to $\mathbf{S}^n$ at $\mathbf{p}$. See Figure 3.36. Tangent vectors will be defined in Chapter 8. For now, aside from the intuitive meaning, take this to mean that $\mathbf{u}$ and $\mathbf{v}$ are tangent to circles $\mathbf{C_u}$ and $\mathbf{C_v}$, respectively, in $\mathbf{S}^n$ through $\mathbf{p}$ and that "tangent at a point $\mathbf{p}$ of a circle with center $\mathbf{c}$" means a vector in the plane containing the circle that is orthogonal to the vector $\mathbf{cp}$. Let $\mathbf{p'}$, $\mathbf{C_u'}$, and $\mathbf{C_v'}$ in $\mathbf{R}^n$ be the images of $\mathbf{p}$, $\mathbf{C_u}$, and $\mathbf{C_v}$, respectively, under the stereographic projection. The vectors $\mathbf{u}$ and $\mathbf{v}$ induce an orientation of the circles $\mathbf{C_u}$ and $\mathbf{C_v}$ (think of $\mathbf{u}$ and $\mathbf{v}$ as velocity vectors of someone walking along the circles) and these orientations induce orientations of the circles $\mathbf{C_u'}$ and $\mathbf{C_v'}$ via the stereographic projection. Choose tangent vectors $\mathbf{u'}$ and $\mathbf{v'}$ to the circles $\mathbf{C_u'}$ and $\mathbf{C_v'}$ at $\mathbf{p'}$ that match their orientation. Let $\theta$ be the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$ and $\theta'$ the angle between $\mathbf{u'}$ and $\mathbf{v'}$.

**3.10.3. Theorem.** The stereographic projection is an *angle-preserving* or *conformal* map, that is, $\theta = \theta'$.

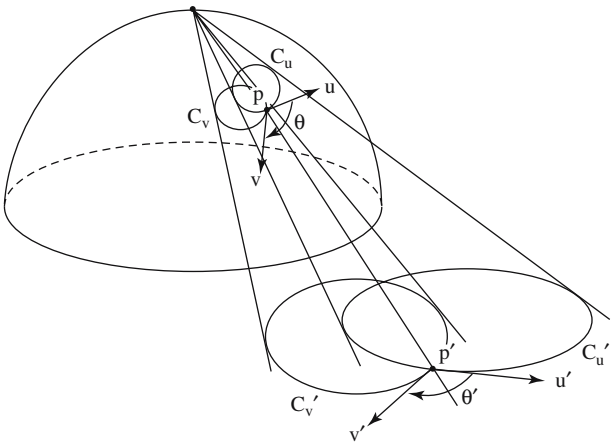



**Figure 3.36.** The stereographic projection preserves angles.

**Proof.** See [HilC99].

**3.10.4. Corollary.** All sphere-preserving maps of $\mathbf{R}^n_\infty$ are angle preserving.

**Proof.** This is an immediate consequence of Theorems 3.10.2 and 3.10.3.

Now let $\mathbf{X}$ be a (n – 1)-sphere in $\mathbf{R}^n$ with center $\mathbf{c}$ and radius r. Let $\mathbf{X'}$ be the sphere in $\mathbf{S}^n$ that is mapped onto $\mathbf{X}$ by the stereographic projection. Choose a point $\mathbf{p}$ on $\mathbf{X}$ and let $\mathbf{p'}$ be the point of $\mathbf{S}^n$ (in $\mathbf{X'}$) that maps onto $\mathbf{p}$. Let σ be the rotation of $\mathbf{S}^n$ around the great circle through $\mathbf{p'}$ and $\mathbf{e}_{n+1}$ that maps $\mathbf{p'}$ to $\mathbf{e}_{n+1}$. Then $\mathbf{Y'} = \sigma(\mathbf{X'})$ is a sphere through $\mathbf{e}_{n+1}$ and its projection to $\mathbf{R}^n$ is a plane $\mathbf{Y}$. Let R be the reflection of $\mathbf{R}^n$ about $\mathbf{Y}$.

**Definition.** The map

$$\mu = p_n(\sigma^{-1})Rp_n(\sigma): \ \mathbf{R}^n_\infty \to \mathbf{R}^n_\infty$$

is called an *n-dimensional inversion* of $\mathbf{R}^n_\infty$ with respect to the sphere $\mathbf{X}$, or simply an *inversion in a sphere*.

**3.10.5. Theorem.** The map μ is defined analytically as follows:

(1) $\mu(\mathbf{c}) = \infty$.
(2) Let $\mathbf{p} \in \mathbf{R}^n$, $\mathbf{p} \neq \mathbf{c}$. Let $\mathbf{q}$ be the point on $\mathbf{X}$ where the ray $\mathbf{Z}$ from $\mathbf{c}$ through $\mathbf{p}$ intersects $\mathbf{X}$. Then $\mu(\mathbf{p})$ is that unique point $\mathbf{z}$ on $\mathbf{Z}$ defined by the equation

$$|\mathbf{cp}||\mathbf{cz}| = |\mathbf{cq}|^2 = r^2.$$

**Proof.** See Figure 3.37. For a proof see [HilC99].

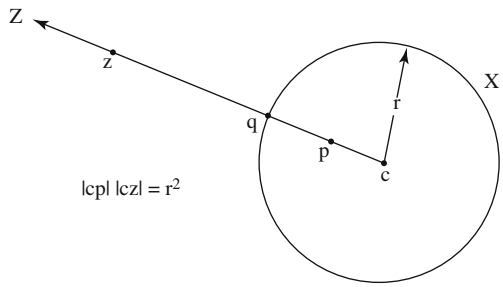



**Figure 3.37.** An inversion of a sphere maps **p** to **z**.

**3.10.6. Theorem.** Every circle-preserving map of $\mathbf{R}^n_\infty$ is the composite of at most three inversions.

**Proof.** See [HilC99].

Finally, there is an interesting connection between the stereographic projection and Poincaré's model of the hyperbolic plane. To learn about this we again refer the reader to [HilC99]. Recall that one of the big developments in geometry in the 19th century was the discovery of non-Euclidean geometry. The big issue was whether the axiom of parallels was a consequence of the other axioms of Euclidean geometry. The *axiom of parallels* asserts that given a line and a point not on the line, there is a unique line through the point that is parallel to the line, that is, does not intersect it. This axiom does not hold in other geometries. In the plane of elliptic geometry there is no parallel line because all lines intersect. In hyperbolic geometry there are an infinite number of lines through a point that are parallel to a given line.

## 3.11 Exercises

### Section 3.4

3.4.1. Let $\ell$ be an ordinary line in $\mathbf{R}^2$. Carefully prove that the set

$$\mathbf{L} = \ell \cup \{\ell^\infty\}$$

in $\mathbf{P}^2$ is in fact a line in $\mathbf{P}^2$.

### Section 3.4.1

3.4.1.1. Find the equation of the line in $\mathbf{P}^2$ through the points [2,–3,1] and [1,0,1].

3.4.1.2. Find the intersection of the lines

$$-X + 2Y - Z = 0 \quad \text{and} \quad 2X + Z = 0$$

in $\mathbf{P}^2$.

3.4.1.3. Find the cross-ratio of the points [1,0,1], [0,1,1], [2,–1,3], and [3,1,2] in $\mathbf{P}^2$.

3.4.1.4. The points $\mathbf{I} = [0,1,-1]$, $\mathbf{O} = [1,0,-2]$, and $\mathbf{U} = [2,-4,0]$ belong to a line $\mathbf{L}$ in $\mathbf{P}^2$.

(a) Find the coordinates of the point [1,2,–4] on $\mathbf{L}$ with respect to $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{U}$.
(b) Find the coordinates of the point [1,2,–4] with respect to $\mathbf{I}' = [1,0,-2]$, $\mathbf{O}' = [0,1,-1]$, and $\mathbf{U}' = [1,1,-3]$.
(c) Find the transformation φ that maps the coordinates with respect to $\mathbf{I}$, $\mathbf{O}$, and $\mathbf{U}$ to the coordinates with respect to $\mathbf{I}'$, $\mathbf{O}'$, and $\mathbf{U}'$.

3.4.1.5. Consider the points $\mathbf{I} = [0,1,-1]$, $\mathbf{J} = [1,0,1]$, $\mathbf{O} = [1,0,-2]$, and $\mathbf{U} = [2,-4,0]$ in $\mathbf{P}^2$.

(a) Find the coordinates of the point [1,2,–4] with respect to $\mathbf{I}$, $\mathbf{J}$, $\mathbf{O}$, and $\mathbf{U}$.
(b) Find the coordinates of the point [1,2,–4] with respect to $\mathbf{I}' = [1,0,-2]$, $\mathbf{J}' = [3,1,1]$, $\mathbf{O}' = [0,1,-1]$, and $\mathbf{U}' = [1,1,-3]$.
(c) Find the transformation φ that maps the coordinates with respect to $\mathbf{I}$, $\mathbf{J}$, $\mathbf{O}$, and $\mathbf{U}$ to the coordinates with respect to $\mathbf{I}'$, $\mathbf{J}'$, $\mathbf{O}'$, and $\mathbf{U}'$.

**Section 3.4.3**

3.4.3.1. Let T be the central projection that projects $\mathbf{R}^2$ onto the line $\mathbf{L}$ defined by $2x - 3y + 6 = 0$ from the point $\mathbf{p} = (5,1)$.

(a) Find the equation for T in two ways:

(1) Using homogeneous coordinates and projective transformations
(2) Finding the intersection of lines from $\mathbf{p}$ with $\mathbf{L}$

(b) Find T(7,1) and T(3,4).

**Section 3.5.1**

3.5.1.1. Let T be the central projection that projects $\mathbf{R}^3$ onto the plane $\mathbf{X}$ defined by $x + y + z = 1$ from the point $\mathbf{p} = (-1,0,0)$.

(a) Find the equations for T in three ways:

(1) Using the usual composites of rigid motions and central projections and homogeneous coordinates
(2) Via the method of frames
(3) Finding the intersections of lines through $\mathbf{p}$ with the plane

(b) Find T(9,0,0) and T(4,0,5).

**Section 3.6**

3.6.1. Consider the conic defined by the equation

$$31x^2 - 10\sqrt{3}xy + 21y^2 + (10\sqrt{3} - 124)x + (20\sqrt{3} - 42)y - 20\sqrt{3} + 1 = 0.$$

(a) Is the conic an ellipse, hyperbola, or parabola?
(b) Find its natural coordinate system.
(c) Determine its focus and directrix.

**Section 3.6.1**

3.6.1.1. Find the projective transformation (like in Example 3.6.1.2) that transforms the conic

$$xy + 2y^2 - x + y + 3 = 0$$

into the unit circle.

3.6.1.2. Find the tangent line to the conic in Exercise 3.6.1.1 at the point (0,1).

3.6.1.3. Find the equation of the conic through the points $\mathbf{p}_1 = (1,1)$, $\mathbf{p}_2 = (2,1+(3/2)\sqrt{3})$, $\mathbf{p}_3 = (5,1)$, $\mathbf{p}_4 = (4,1-(3/2)\sqrt{3})$ and that has tangent line $x - 1 = 0$ at the point $\mathbf{p}_1$.

3.6.1.4. Find the equation of the conic through the points $\mathbf{p}_1 = (1,2)$, $\mathbf{p}_2 = (-3,2)$, and $\mathbf{p}_3 = (-1,1)$ and which has tangent lines $y - x - 1 = 0$ and $x + y + 1 = 0$ at the point $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively.

3.6.1.5. Find the equation of the conic through the points $\mathbf{p}_1 = (2,-1)$ and $\mathbf{p}_2 = (4,-2)$ that has tangent lines $y = -1$ and $x + y - 2 = 0$ at those points, respectively, and is also tangent to the line $2x - y - 1 = 0$.

3.6.1.6. Solve conic design problem 4.

3.6.1.7. Solve conic design problem 5.

**Section 3.7**

3.7.1. Consider the following quadric surface

$$x^2 + y^2 + 2z^2 - 2xy - \sqrt{2}x - \sqrt{2}y - 2 = 0.$$

(a) Determine its type.
(b) Find its tangent plane at the point (0,0,1).

**Section 3.10**

3.10.1. Show that the inverse

$$p_n^{-1} : \mathbf{R}^n \to \mathbf{S}^n - \{\mathbf{e}_{n+1}\}$$

of the stereographic projection is defined by

$$p_n^{-1}(\mathbf{y}) = \frac{1}{|\mathbf{y}|^2 + 1}\left(2y_1, 2y_2, \ldots, 2y_n, |\mathbf{y}|^2 - 1\right), \mathbf{y} \in \mathbf{R}^n.$$

CHAPTER 4

# Advanced Calculus Topics

## 4.1 Introduction

The object of this chapter is to introduce basic topological concepts as they apply to $\mathbf{R}^n$ and to cover some important topics in advanced calculus. The reader is assumed to have had the basic three-semester sequence of calculus and it is not our intent to redo that material here. Our emphasis will be on multivariable functions and their properties, the assumption being that the reader has a reasonable understanding of functions of a single variable. Proofs are given in those cases where it was thought to be helpful in understanding some new ideas or if they involved some geometric insights.

Section 4.2 introduces the topological concepts. We limit the discussion to those that are specifically needed for advanced calculus and leave the more general study of topology to Chapter 5. Section 4.3 describes the derivative of vector-valued functions of several variables and related results that generalize well-known properties of real-valued functions of a single variable. The inverse function theorem and the implicit function theorem are discussed in Section 4.4. These are such important theorems and get used so often that we give a fairly detailed outline of their proofs. Many results in differential topology and algebraic geometry would be impossible without them. Next, Section 4.5 develops the basic results regarding critical points of functions and this leads to Morse theory, which is described in Section 4.6. The problem of finding zeros of functions is addressed in Section 4.7. Section 4.8 reviews basic facts about integrating functions of several variables. Finally, in Sections 4.9 and 4.9.1, we start a brief overview of the topic of differential forms and their integrals that will be continued in Section 8.12.

## 4.2 The Topology of Euclidean Space

Topology is the study of the most basic properties of point sets such as, what is meant by a neighborhood of a point, what open and closed sets are, and what makes a func-

tion continuous. It is the foundation of calculus, analysis, and any sort of geometric investigations. In this section we look only at some of the important definitions as they apply to Euclidean space. Chapter 5 will study abstract topological spaces. Because we shall encounter many of the same definitions there in a more general context, we shall postpone some proofs and results to that chapter to avoid stating and proving theorems twice. By and large, this section is simply a collection of definitions and essentially immediate consequences. By carefully asking "what does this mean?" the reader should have little difficulty in proving most theorems.

Unless stated otherwise, all points and sets in this section belong to $\mathbf{R}^n$ for some fixed but arbitrary n.

**Definition.** A set **N** is said to be a *neighborhood* of a point **p** if it contains an open ball about **p**, that is, there exists an $\varepsilon > 0$ such that $\mathbf{B}^n(\mathbf{p},\varepsilon) \subseteq \mathbf{N}$.

The important part of the neighborhood **N** is that it contains some open ball about **p**. We do not care whether it contains some other "junk." For example, the open interval (–1,1) together with the set {–100,23,5} would be called a neighborhood of the origin in **R**. Note also that this definition and many others depend on the dimension of the Euclidean space with respect to which we are making the definition. For instance, the interval (–1,1) is a neighborhood of the origin for **R** but **not** for $\mathbf{R}^2$. The open unit ball $\mathbf{B}^2$, on the other hand, **is** a neighborhood of the origin for $\mathbf{R}^2$.

**Definition.** A subset **X** of $\mathbf{R}^n$ is called an *open set* if for all points **p** in **X** there is an $\varepsilon > 0$ such that $\mathbf{B}^n(\mathbf{p},\varepsilon) \subseteq \mathbf{X}$. **X** is said to be a *closed set* if the complement of **X**, $\mathbf{R}^n - \mathbf{X}$, is open.

Specifying the open sets of a set is what defines its "topology." (A precise definition of the term "topology" is given in Chapter 5). "Open" and "closed" are dual concepts. The open interval (0,1) is an open set in **R** and the closed interval [0,1] is a closed set. Our definitions are compatible with the old usage of the terms. A set does not have to be either open or closed. For example, the "half-open" ("half-closed") interval (0,1] is neither. The set $\mathbf{R}^n$ is **both** open and closed as is the empty set $\phi$ but these are the only subsets of $\mathbf{R}^n$ that are both open and closed. A single point is always a closed set. In practice, sets tend to be open if the "<" relation is used in their definition and closed if the "≤" relation is used.

**4.2.1. Proposition**

(1) The arbitrary union of open sets is an open set.
(2) A finite intersection of open sets is open.
(3) The arbitrary intersection of closed sets is a closed set.
(4) A finite union of closed sets is closed.

**Proof.** Part (1) is easy. To prove (2), let $\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_k$ be open sets and let **p** be a point in their intersection. Since each $\mathbf{O}_i$ is an open set, there exists an $\varepsilon_i > 0$, so that $\mathbf{B}^n(\mathbf{p},\varepsilon_i) \subseteq \mathbf{O}_i$. If

$$\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k\},$$

then the open ball $\mathbf{B}^n(\mathbf{p},\varepsilon)$ is contained in all the sets $\mathbf{O}_i$ and hence in their intersection, proving (2). Parts (3) and (4) follow from parts (1) and (2) and the identity

$$\bigcap_{i=1}^{k}(\mathbf{R}^n - \mathbf{O}_i) = \mathbf{R}^n - \bigcup_{i=1}^{k}\mathbf{O}_i.$$

The identities

$$\bigcap_{n=1}^{\infty}\left(-\frac{1}{n},\frac{1}{n}\right) = 0 \quad \text{and} \quad \bigcup_{n=2}^{\infty}\left[\frac{1}{n},1-\frac{1}{n}\right] = (0,1)$$

show that arbitrary intersections of open sets need not be open and arbitrary unions of closed sets need not be closed.

**Definition.** A point $\mathbf{p}$ is a *limit* or *accumulation point* of a set $\mathbf{X}$ if every neighborhood of $\mathbf{p}$ meets $\mathbf{X}$ in a point other than $\mathbf{p}$. More precisely, for every $\varepsilon > 0$,

$$(\mathbf{B}^n(\mathbf{p},\varepsilon) - \{\mathbf{p}\}) \cap \mathbf{X} \neq \phi.$$

An *isolated point* of $\mathbf{X}$ is a point of $\mathbf{X}$ that has a neighborhood containing no point of $\mathbf{X}$ other than itself (that is, it is not a limit point).

For example, 1 is a limit point of (0,1), but 1.0001 is not because $\mathbf{B}^1(1.0001,0.00005)$ is disjoint from (0,1). A point $\mathbf{p}$ in the set $\mathbf{X}$ is **not** necessarily a limit point of $\mathbf{X}$. Only those points whose neighborhoods contain infinitely many points of $\mathbf{X}$ are. For example, if $\mathbf{X} = [0,1] \cup \{2\}$, then 2 is an isolated point and **not** a limit point of $\mathbf{X}$, but every point in [0,1] **is** a limit point of $\mathbf{X}$.

**Definition.** The *closure* of a set $\mathbf{X}$, denoted by cl($\mathbf{X}$), is defined by

$$\mathrm{cl}(\mathbf{X}) = \mathbf{X} \cup \{\mathbf{p} \mid \mathbf{p} \text{ is a limit point of } \mathbf{X}\}$$

For example,

$$\mathrm{cl}((0,1)) = [0,1] \quad \text{and} \quad \mathrm{cl}([0,1]) = [0,1].$$

The closure of the set of rationals in the reals is the reals.

**4.2.2. Proposition**

(1) For every set $\mathbf{X}$, cl($\mathbf{X}$) is a closed set.
(2) If $\mathbf{X}$ is closed, then cl($\mathbf{X}$) = $\mathbf{X}$.

**Proof.** We prove (1) and leave (2) as an exercise (Exercise 4.2.4). We need to show that

$$\mathbf{Y} = \mathbf{R}^n - \mathrm{cl}(\mathbf{X})$$

is an open set. Let $\mathbf{p} \in \mathbf{Y}$. See Figure 4.1. Since $\mathbf{p} \notin \mathrm{cl}(\mathbf{X})$, there is an $\varepsilon > 0$, so that $\mathbf{B}^n(\mathbf{p},\varepsilon) \cap \mathbf{X} = \phi$. To show that $\mathbf{B}^n(\mathbf{p},\varepsilon)$ does not contain any limit point of $\mathbf{X}$, let $\mathbf{q} \in \mathbf{B}^n(\mathbf{p},\varepsilon)$ and let $r = \varepsilon - |\mathbf{pq}|$. Then

$$\mathbf{B}^n(\mathbf{q},r) \subseteq \mathbf{B}^n(\mathbf{p},\varepsilon) \subseteq \mathbf{R}^n - \mathbf{X},$$

so that $\mathbf{q}$ is not a limit point of $\mathbf{X}$.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$. The *boundary* of $\mathbf{X}$, denoted by bdry($\mathbf{X}$), is defined by

$$\mathrm{bdry}(\mathbf{X}) = \{\mathbf{p} \mid \text{every neighborhood } \mathbf{N} \text{ of } \mathbf{p} \text{ meets both } \mathbf{X} \text{ and its complement, that is, } \mathbf{N} \cap \mathbf{X} \neq \phi \text{ and } \mathbf{N} \cap (\mathbf{R}^n - \mathbf{X}) \neq \phi\}.$$

The *interior* of $\mathbf{X}$, denoted by int($\mathbf{X}$), is defined by

$$\mathrm{int}(\mathbf{X}) = \{\mathbf{p} \in \mathbf{X} \mid \mathbf{p} \text{ has a neighborhood } \mathbf{N} \text{ with } \mathbf{N} \subseteq \mathbf{X}\}.$$

See Figure 4.2. For example, if n = 1, then

$$\mathrm{bdry}((0,1)) = \{0,1\} \quad \text{and} \quad \mathrm{int}([0,1]) = (0,1).$$

Note that boundary points of a set do not necessarily belong to the set. The boundary of the set of rationals in the reals is all of the reals and the interior of this set is empty. For "nice" sets the boundary and the interior are the obvious sets; however,

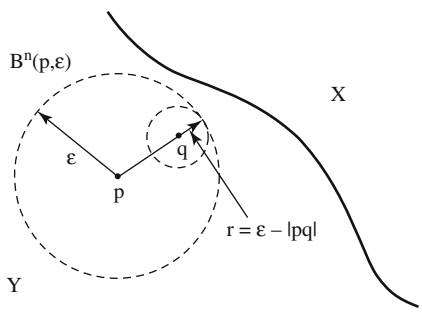



**Figure 4.1.** Proving that cl(**X**) is a closed set.

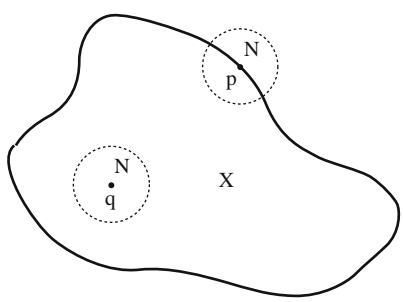



**Figure 4.2.** A boundary point **p** and an interior point **q** of a set **X**.

Exercise 4.2.6 shows that the definitions may have some unexpected consequences in certain cases.

The boundary and interior of a set change if we change the containing space $\mathbf{R}^n$. For example, the boundary of [0,1] in $\mathbf{R}^2$ is all of [0,1] and its interior is the empty set.

**4.2.3. Proposition**

(1) The boundary of a set is a closed set.
(2) If $\mathbf{X} \subseteq \mathbf{R}^n$, then $\text{bdry}(\mathbf{R}^n - \mathbf{X}) = \text{bdry}(\mathbf{X})$.
(3) A set $\mathbf{X}$ is closed if and only if $\text{bdry}(\mathbf{X}) \subseteq \mathbf{X}$.
(4) The interior of a set is an open set.
(5) $\text{int}(\mathbf{X}) = \mathbf{X} - \text{bdry}(\mathbf{X})$.

**Proof.** Easy.

**Definition.** A subset $\mathbf{X}$ of $\mathbf{R}^n$ is called a *bounded set* if $\mathbf{X} \subseteq \mathbf{B}^n(r)$ for some $r > 0$. If $\mathbf{X}$ is not bounded, then it is said to be *unbounded*.

For example, the set of integers is an unbounded set in $\mathbf{R}$. Of course, the whole set $\mathbf{R}$ is an unbounded subset of $\mathbf{R}$. The interval [–1,10] is a bounded set because it is contained in $(-50,50) = \mathbf{B}^1(50)$.

**Definition.** Let $\mathbf{X}$ be a subset of $\mathbf{R}^n$. If $S = \{\mathbf{U}_\alpha\}$ is a collection of subsets of $\mathbf{R}^n$ whose union contains $\mathbf{X}$, then S is called a *cover* of $\mathbf{X}$. If all the $\mathbf{U}_\alpha$ are open sets in $\mathbf{R}^n$, then S is called an *open cover* of $\mathbf{X}$. If all the $\mathbf{U}_\alpha$ are closed sets in $\mathbf{R}^n$, then S is called a *closed cover* of $\mathbf{X}$.

For example, $\{(1/n, 1 - 1/n) \mid n = 2, 3, \ldots\}$ is an open cover of the subset (0,1) of $\mathbf{R}$.

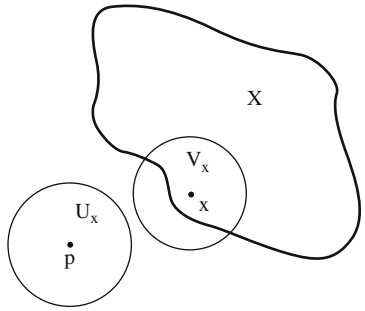
**Definition.** A subset $\mathbf{X}$ of $\mathbf{R}^n$ is said to be *compact* if every open cover of $\mathbf{X}$ has a *finite subcover*, that is, there is a finite collection of sets from the cover that already cover $\mathbf{X}$.

**4.2.4. Theorem.** (The Heine-Borel-Lebesque Theorem) A subset $\mathbf{X}$ of $\mathbf{R}^n$ is compact if and only if it is closed and bounded.

**Proof.** We shall prove half of the theorem, namely, that a compact set is closed and bounded, and leave the converse to the next chapter (Theorem 5.5.6). One reason for proving at least part of the theorem here is to show how the property of covers having finite subcovers gets used. Basically, when one has a finite collection of objects, e.g., numbers, then one can talk about the smallest or largest. This is not possible with infinite collections.

Assume that $\mathbf{X}$ is a compact subset of $\mathbf{R}^n$. To prove that $\mathbf{X}$ is closed we need to show that $\mathbf{R}^n - \mathbf{X}$ is open. Let $\mathbf{p} \in \mathbf{R}^n - \mathbf{X}$. For every $\mathbf{x} \in \mathbf{X}$ there is a ball neighborhood $\mathbf{U}_\mathbf{x}$ and $\mathbf{V}_\mathbf{x}$ of $\mathbf{p}$ and $\mathbf{x}$, respectively, such that $\mathbf{U}_\mathbf{x} \cap \mathbf{V}_\mathbf{x} = \phi$. See Figure 4.3. The

**Figure 4.3.** Proving that compact sets are closed.




collection $\{\mathbf{V_x}\}$ is an open cover of $\mathbf{X}$. Since $\mathbf{X}$ is compact, there is a finite subcover $\{\mathbf{V_{x_i}}\}_{1\le i\le k}$. It follows that

$$\mathbf{U} = \bigcap_{i=1}^{k} \mathbf{U_{x_i}}$$

is an open neighborhood of $\mathbf{p}$ contained in $\mathbf{R}^n - \mathbf{X}$ and so $\mathbf{R}^n - \mathbf{X}$ is open.

Next, we show that $\mathbf{X}$ is bounded. For every $\mathbf{x} \in \mathbf{X}$, the ball $\mathbf{B_x} = \mathbf{B}^n(|\mathbf{x}| + 1)$ contains $\mathbf{x}$ in its interior. Let $\{\mathbf{B_{x_i}}\}_{1\le i\le k}$ be a finite subcover of the open cover $\{\mathbf{B_x}\}$. If $r = \max_{1\le i\le k} \{|\mathbf{x}_i| + 1\}$, then $\mathbf{X} \subseteq \mathbf{B}^n(r)$.

The Heine-Borel-Lebesque theorem is one of the fundamental theorems in analysis because it abstracts an essential property of sets that is needed to make many results dealing with continuous functions valid. Although the "correct" definition of compactness is in terms of open covers of sets, the theorem is often used to justify defining a set to be compact if it is closed and bounded. This definition is certainly easier to understand if one is not very familiar with analysis. Since the concepts are equivalent, it does not matter much from a practical point of view. At any rate, it follows that all closed intervals [a,b], all disks $\mathbf{D}^n(\mathbf{p},r)$, and all spheres $\mathbf{S}^n$ are compact.

The definitions above were all relative to a fixed Euclidean space $\mathbf{R}^n$. For example, if we were to say that a set is open, then we really are, more accurately, saying that it is open **in** $\mathbf{R}^n$. It is important to have a version of these definitions that is relative to other spaces besides $\mathbf{R}^n$. Specifically, we want to be able to talk about sets that are open **in** $\mathbf{X}$ with respect to any other given space $\mathbf{X}$.

**Definition.** A subset $\mathbf{A}$ of a set $\mathbf{X}$ in $\mathbf{R}^n$ is said to be *open in* $\boldsymbol{X}$, or *relatively open*, if there is an open set $\mathbf{O}$ **in** $\mathbf{R}^n$ with the property that $\mathbf{A} = \mathbf{O} \cap \mathbf{X}$. $\mathbf{A}$ is said to be *closed in* $\boldsymbol{X}$, or *relatively closed*, if $\mathbf{X} - \mathbf{A}$ is open in $\mathbf{X}$.

For example, the subset

$$\mathbf{A} = \left\{(\cos\theta, \sin\theta) \,\middle|\, 0 < \theta < \frac{\pi}{2}\right\}$$

of the plane is **not** open in the plane, but it **is** open in the unit circle $\mathbf{S}^1$ because $\mathbf{A}$ is the intersection of $\mathbf{S}^1$ with the open set

$$\mathbf{O} = \left\{ (r\cos\theta, r\sin\theta) \,\middle|\, \frac{1}{2} < r < 2 \quad \text{and} \quad 0 < \theta < \frac{\pi}{2} \right\}.$$

See Figure 4.4.

It is easy to show that the relative concepts of open and closed do not depend on $\mathbf{R}^n$. Specifically, if $\mathbf{X}$ is also a subset of $\mathbf{R}^m$, then $\mathbf{A}$ will be open/closed in $\mathbf{X}$ with respect to $\mathbf{R}^n$ if and only if it is open/closed with respect to $\mathbf{R}^m$. Although we shall not give the formal definitions, we now also get the obvious relative concepts of limit point of $\mathbf{A}$ *in* $\boldsymbol{X}$, closure of $\mathbf{A}$ *in* $\boldsymbol{X}$, and boundary of $\mathbf{A}$ *in* $\boldsymbol{X}$.

Next, we turn to the topology of continuity. Invariably, whenever one defines some structure in mathematics it is useful to define maps ("morphisms") that preserve this structure. Since the only topological structure that we have on $\mathbf{R}^n$ at the moment is that of open sets, it is natural to define these maps in terms of them. We begin with a notion of limits.

**Definition.** We say that *a sequence of points* $\mathbf{p}_1, \mathbf{p}_2, \ldots$ *has a limit point* $\mathbf{p}$, and write

$$\lim_{i\to\infty} \mathbf{p}_i = \mathbf{p},$$

if for every $\varepsilon > 0$ there is an N so that $i > N$ implies that

$$|\mathbf{p}_i - \mathbf{p}| < \varepsilon.$$

When it comes to functions, it is assumed that the reader has seen the usual $\varepsilon$-$\delta$ type definition for continuity, at least in the case of real-valued functions of one variable. That definition extends to functions of more variables almost verbatim.

Let $\mathbf{X} \subseteq \mathbf{R}^m$ and $\mathbf{Y} \subseteq \mathbf{R}^n$.

**Definition.** Let $f: \mathbf{X} \to \mathbf{Y}$ and let $\mathbf{p} \in \mathbf{X}$. We say that *the function f has a limit* $\boldsymbol{L}$ *at* $\boldsymbol{p}$ and write

$$\lim_{\mathbf{q}\to\mathbf{p}} f(\mathbf{q}) = \mathbf{L},$$

if $\mathbf{L} \in \mathbf{Y}$ and for every $\varepsilon > 0$ there is a $\delta > 0$ so that $0 < |\mathbf{q} - \mathbf{p}| < \delta$ and $\mathbf{q} \in \mathbf{X}$ implies that

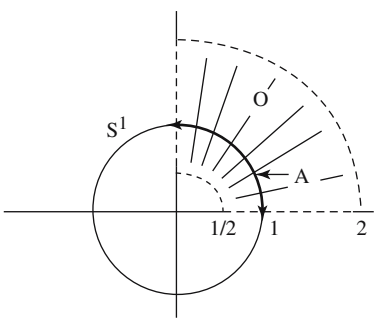



**Figure 4.4.** An open set in the circle.

$$|f(\mathbf{q}) - \mathbf{L}| < \varepsilon.$$

The function is said to be *continuous* at $\mathbf{p}$ if

$$\lim_{\mathbf{q}\to\mathbf{p}} f(\mathbf{q}) = f(\mathbf{p}).$$

This leads to the classical definition of a continuous function.

**First definition of a continuous function.** The function $f:\mathbf{X} \to \mathbf{Y}$ is said to be *continuous* if it is continuous at every point of $\mathbf{X}$.

There are conceptually better and cleaner definitions. The motivation for our new definitions lies in the fact that one can rewrite the definition of continuity of a function f at a point $\mathbf{p}$ as follows:

$$\text{For every } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ so that } f\big(\mathbf{B}^m(\mathbf{p},\delta) \cap \mathbf{X}\big) \subseteq \mathbf{B}^n(f(\mathbf{p}),\varepsilon) \cap \mathbf{Y}. \tag{4.1}$$

**Second definition of a continuous function.** A function $f:\mathbf{X} \to \mathbf{Y}$ is said to be *continuous* if $f^{-1}(\mathbf{V})$ is open in $\mathbf{X}$ for every open set $\mathbf{V}$ in $\mathbf{Y}$.

**Third definition of a continuous function.** A function $f:\mathbf{X} \to \mathbf{Y}$ is said to be *continuous* if $f^{-1}(\mathbf{V})$ is closed in $\mathbf{X}$ for every closed set $\mathbf{V}$ in $\mathbf{Y}$.

The second and third definitions are the "right" definitions, which extend to abstract topological spaces as will be seen in Chapter 5. The second is actually the most common. Since the next theorem shows that the three definitions are in fact equivalent, we shall not distinguish between them.

**4.2.5. Theorem.** The three definitions of continuity are equivalent.

**Proof.** We show that the first and second definitions are equivalent and leave the rest as an exercise for the reader. Assume that f is continuous with respect to the first definition. Let $\mathbf{V}$ be an open set in $\mathbf{Y}$. We need to show that $\mathbf{U} = f^{-1}(\mathbf{V})$ is open in $\mathbf{X}$. See Figure 4.5. Let $\mathbf{p}$ be any point in $\mathbf{U}$ and let $\mathbf{q} = f(\mathbf{p})$. Choose $\varepsilon > 0$ so that $\mathbf{O_q} = \mathbf{B}^n(\mathbf{q},\varepsilon) \cap \mathbf{Y} \subseteq \mathbf{V}$. By statement (4.1), there is a $\delta > 0$ so that $f(\mathbf{O_p}) \subseteq \mathbf{O_q}$, where $\mathbf{O_p} = \mathbf{B}^m(\mathbf{p},\delta) \cap \mathbf{X}$. Therefore, $\mathbf{p}$ belongs to an open subset $\mathbf{O_p}$ of $\mathbf{X}$ that is contained in $\mathbf{U}$. Since $\mathbf{p}$ was an arbitrary point of $\mathbf{U}$, we have shown that $\mathbf{U}$ is an open set in $\mathbf{X}$.

Conversely, assume that f is continuous with respect to the second definition of continuity. Let $\mathbf{p} \in \mathbf{X}$ and set $\mathbf{q} = f(\mathbf{p})$. Let $\varepsilon > 0$. Then $\mathbf{O_q} = \mathbf{B}^n(\mathbf{q},\varepsilon) \cap \mathbf{Y}$ is an open set

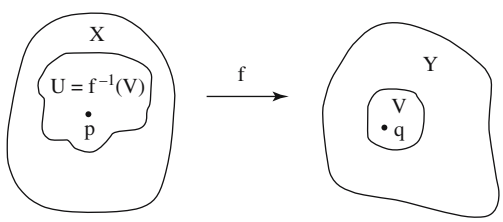



**Figure 4.5.** Understanding continuous functions.

in $\mathbf{Y}$ and $\mathbf{U} = f^{-1}(\mathbf{O_q})$ is open in $\mathbf{X}$. Therefore, there is a $\delta > 0$, so that $\mathbf{O_p} = \mathbf{B}^m(\mathbf{p},\delta) \cap \mathbf{X} \subseteq \mathbf{U}$. Clearly, $f(\mathbf{O_p}) \subseteq \mathbf{O_q}$. In other words, f is continuous at $\mathbf{p}$. We are done.

**4.2.6. Proposition.** The composition of continuous maps is continuous.

**Proof.** The proof is easy using the second or third definition of continuity.

**Definition.** If a function $f: \mathbf{X} \to \mathbf{Y}$ is not continuous at a point $\mathbf{p}$ but

$$\mathbf{L} = \lim_{\mathbf{q}\to\mathbf{p}} f(\mathbf{q})$$

exists, then the point $\mathbf{p}$ is called a *removable discontinuity* of the function f (because we could make f continuous at $\mathbf{p}$ simply by redefining f to equal $\mathbf{L}$ at $\mathbf{p}$). Any other point $\mathbf{p}$ where f is discontinuous is called an *essential discontinuity*.

For example, consider the functions $f, g: \mathbf{R} \to \mathbf{R}$ defined by

$$\begin{aligned} &f(x) = 1, && \text{for } x \neq 0, \quad f(0) = 2, \quad \text{and} \\ &g(x) = \sin(1/x), && \text{for } x \neq 0, \quad g(0) = 2. \end{aligned}$$

The function f has a removable discontinuity at 0, whereas the function g has an essential discontinuity there.

If a function $f: \mathbf{X} \to \mathbf{Y}$ is continuous it is continuous at every point. If $\varepsilon > 0$, then the first definition of continuity gives us a $\delta > 0$ so that points within $\delta$ of a point $\mathbf{p}$ will get mapped to a point within $\varepsilon$ of $f(\mathbf{p})$. However, it is important to realize that the $\delta$ depends on $\mathbf{p}$. It may **change** from point to point. A nice situation is one where one can choose a $\delta$ that will work for **all** points.

**Definition.** A function $f: \mathbf{X} \to \mathbf{Y}$ is said to be *uniformly continuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ so that for all $\mathbf{p}, \mathbf{q} \in \mathbf{X}$ with $|\mathbf{q} - \mathbf{p}| < \delta$ we have that $|f(\mathbf{q}) - f(\mathbf{p})| < \varepsilon$.

Uniform continuity is a very important concept in analysis.

**4.2.7. Example.** It is not hard to show that the function

$$f: (0,1] \to \mathbf{R}, \quad f(x) = 1/x,$$

is continuous but **not** uniformly continuous. For a fixed $\varepsilon$, the $\delta$ that works at a point x gets smaller and smaller as x approaches 0. The core problem is that (0,1] is not compact.

**4.2.8. Theorem.** If $f: \mathbf{X} \to \mathbf{Y}$ is a continuous function and if $\mathbf{X}$ is compact, then f is uniformly continuous.

**Proof.** See [Buck78] or [Eise74]. See also Theorem 5.5.13.

The following property is weaker than uniform continuity but is sometimes adequate:

**Definition.** A function $f: \mathbf{X} \to \mathbf{Y}$ is said to satisfy a *Lipschitz condition* on $\mathbf{X}$ if there is constant $M > 0$, so that $|f(\mathbf{q}) - f(\mathbf{p})| \le M \mid \mathbf{p} - \mathbf{q}|$ for all $\mathbf{p}, \mathbf{q} \in \mathbf{X}$.

**Definition.** A function $f: \mathbf{X} \to \mathbf{Y}$ is said to be a *homeomorphism* if it is one-to-one and onto and both it **and** its inverse are continuous.

To a topologist all homeomorphic spaces look alike and the main problem is to classify spaces up to homeomorphism.

**Definition.** A set $\mathbf{X}$ is said to be *connected* if $\mathbf{X}$ cannot be written as the union of two subsets $\mathbf{A}$ and $\mathbf{B}$ that are nonempty disjoint open sets in $\mathbf{X}$.

**4.2.9. Theorem.** Every connected subset of $\mathbf{R}$ is either $\mathbf{R}$ itself or an interval of the form $[a,b]$, $(a,b)$, $(a,b]$, $[a,b)$, $(-\infty,a]$, $(-\infty,a)$, $[a,\infty)$, or $(a,\infty)$.

**Proof.** See [Eise74].

A more intuitive notion of connectedness is the following:

**Definition.** A set $\mathbf{X}$ is said to be *path-connected* if for all points $\mathbf{p}$ and $\mathbf{q}$ in $\mathbf{X}$ there is a continuous function $f: [0,1] \to \mathbf{X}$ with $f(0) = \mathbf{p}$ and $f(1) = \mathbf{q}$.

**4.2.10. Proposition.** A path-connected space is connected.

**Proof.** See [Eise74].

The converse of Proposition 4.2.10 is not necessarily true.

**Definition.** A *component* of a set is a maximal connected subset.

Next, we describe some properties that are preserved by continuous maps.

**4.2.11. Theorem.** The continuous image of a compact set is compact.

**Proof.** Let $\mathbf{X}$ be a compact set and let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous onto map. Let $\{\mathbf{O}'_\alpha\}$ be an open cover of $\mathbf{Y}$ and let $\mathbf{O}_\alpha = f^{-1}(\mathbf{O}'_\alpha)$. By continuity of f, each $\mathbf{O}_\alpha$ is open in $\mathbf{X}$. In fact, $\{\mathbf{O}_\alpha\}$ is an open cover of $\mathbf{X}$. Since $\mathbf{X}$ is compact, there is a finite subcover $\{\mathbf{O}_{\alpha_i}\}_{1 \le i \le k}$. Clearly, $\{\mathbf{O}'_{\alpha_i}\}$ is a finite cover of $\mathbf{Y}$ and the theorem is proved. See also Theorem 5.5.8.

**4.2.12. Theorem.** The continuous image of a connected set is connected.

**Proof.** Let $\mathbf{X}$ be a connected set and let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous map which is onto $\mathbf{Y}$. Suppose that $\mathbf{Y}$ is not connected. Then $\mathbf{Y} = \mathbf{O}_1 \cup \mathbf{O}_2$, where $\mathbf{O}_i$ is open in $\mathbf{Y}$ and $\mathbf{O}_1 \cap \mathbf{O}_2 = \phi$. But then $\mathbf{X}$ would be a disjoint union of open sets $f^{-1}(\mathbf{O}_1)$ and $f^{-1}(\mathbf{O}_2)$, which would contradict the connectedness of $\mathbf{X}$. Note that since f is onto, neither $f^{-1}(\mathbf{O}_1)$ nor $f^{-1}(\mathbf{O}_2)$ is empty.

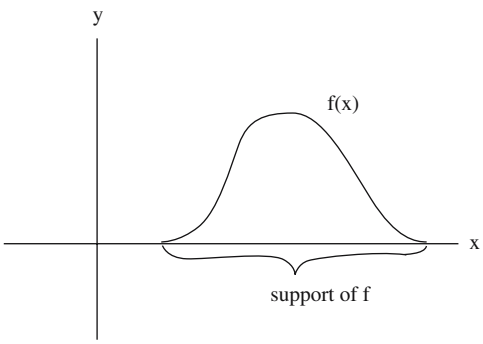



**Figure 4.6.** The support of a function.

**Definition.** The *support* of a function $f: \mathbf{R}^n \to \mathbf{R}$ is the closure of the set of points where f is nonzero. (See Figure 4.6.)

## 4.3 Derivatives

Given a function $f: \mathbf{R} \to \mathbf{R}$, the usual definition of the derivative for f at a point a is

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}, \tag{4.2}$$

if the limit exists. This works fine for functions of one variable, but just like the single number "slope" cannot capture the direction of vectors in dimension larger than two, we need a different definition of derivative in higher dimensions.

**Definition.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$. A function $f: \mathbf{U} \to \mathbf{R}^m$ is said to be *differentiable* at $\mathbf{p} \in \mathbf{U}$ if there is linear transformation $T: \mathbf{R}^n \to \mathbf{R}^m$ such that

$$\lim_{h \to 0} \frac{|f(\mathbf{p}+\mathbf{h}) - f(\mathbf{p}) - T(\mathbf{h})|}{|\mathbf{h}|} = 0. \tag{4.3}$$

In that case, T will be called the *derivative of f at* $\boldsymbol{p}$ and will be denoted by $Df(\mathbf{p})$.

See Figure 4.7. Note that if n = m = 1, then equation (4.3) is simply a rewrite of the equation in (4.2) if we define $T(h) = f'(a)h$. In other words, in arbitrary dimensions, the derivative needs to be replaced by a linear transformation rather than having it simply be a real number. The fact that a linear transformation from the reals to the reals could be specified by a real number obscured what was really going on.

**4.3.1. Proposition.** The linear transformation T in equation (4.3) is unique if it exists.

**Proof.** Suppose that S is another linear transformation satisfying equation (4.3) where T is replaced by S. Then

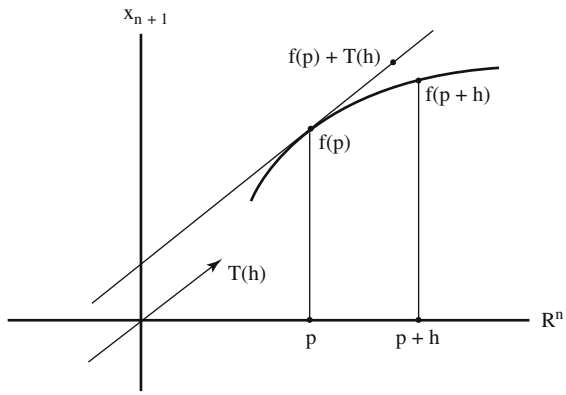



**Figure 4.7.** The geometry of the derivative.

$$\begin{aligned}\lim_{h\to 0}\frac{|S(\mathbf{h})-T(\mathbf{h})|}{|\mathbf{h}|} &= \lim_{h\to 0}\frac{|S(\mathbf{h})-(f(\mathbf{p}+\mathbf{h})-f(\mathbf{p}))+f(\mathbf{p}+\mathbf{h})-f(\mathbf{p})-T(\mathbf{h})|}{|\mathbf{h}|}\\ &\le \lim_{h\to 0}\frac{|S(\mathbf{h})-(f(\mathbf{p}+\mathbf{h})-f(\mathbf{p}))|}{|\mathbf{h}|}+\lim_{h\to 0}\frac{|f(\mathbf{p}+\mathbf{h})-f(\mathbf{p})-T(\mathbf{h})|}{|\mathbf{h}|}\\ &= 0.\end{aligned}$$

Now $t\mathbf{p} \to \mathbf{0}$ as $t \to 0$, for any $\mathbf{p} \in \mathbf{R}^n$. Therefore, if $\mathbf{p} \neq 0$, then we can let $t\mathbf{p}$ play the role of $\mathbf{h}$ above to get

$$0 = \lim_{t\to 0}\frac{|S(t\mathbf{p})-T(t\mathbf{p})|}{|t\mathbf{p}|} = \frac{|S(\mathbf{p})-T(\mathbf{p})|}{|\mathbf{p}|},$$

so that $S(\mathbf{p}) = T(\mathbf{p})$.

One should think of the derivative as the linearization of a function. No **one** linear transformation approximates f. Instead, there are lots of linear transformations, one at every point, which locally approximate f. The linear map $T_{\mathbf{p}}$ defined by

$$T_{\mathbf{p}}(\mathbf{q}) = f(\mathbf{p}) + Df(\mathbf{p})(\mathbf{q}-\mathbf{p}), \tag{4.4}$$

whose graph is a plane, is what approximates the graph of f in a neighborhood of $\mathbf{p}$.

**Definition.** The graph of the linear transformation $T_{\mathbf{p}}: \mathbf{R}^n \to \mathbf{R}^m$ defined by equation (4.4) is called the *tangent plane* to the graph of f at the point $(\mathbf{p}, f(\mathbf{p}))$ in $\mathbf{R}^n \times \mathbf{R}^m$. If $n = m = 1$, then it is usually called the *tangent line*.

The current notion of a tangent plane is rather special. We shall have much more to say about tangent planes and begin to see their importance in Chapter 8.

**4.3.2. Example.** The derivative of $f(x) = x^2$ at a is defined by $Df(a)(x) = 2ax$.

**Proof.** Let

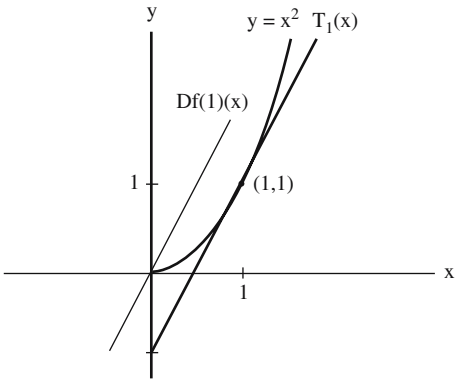



**Figure 4.8.** Example 4.3.2.

$$L = \lim_{h\to 0} \frac{|f(a+h) - f(a) - 2ah|}{|h|}.$$

We must show that the limit L exists and is zero. But

$$L = \lim_{h\to 0} \frac{\left|(a+h)^2 - a^2 - 2ah\right|}{|h|} = \lim_{h\to 0} \frac{h^2}{|h|} = \lim_{h\to 0} |h| = 0.$$

Figure 4.8 shows the graph of the linear map Df(1)(x), which is the derivative of f at 1, and the graph of the linear function $T_1(x)$ defined by equation (4.4), which defines that tangent line to f at (1,1).

**4.3.3. Example.** The derivative of $f(x,y) = x^2$ at (a,b) is defined by $Df(a,b)(x,y) = 2ax$.

**Proof.** We must show that

$$L = \lim_{(\Delta x,\Delta y)\to \mathbf{0}} \frac{|f(a+\Delta x, b+\Delta y) - f(a,b) - 2a\Delta x|}{|(\Delta x, \Delta y)|} = 0.$$

But

$$L = \lim_{(\Delta x,\Delta y)\to \mathbf{0}} \frac{\left|(a+\Delta x)^2 - a^2 - 2a\Delta x\right|}{\sqrt{\Delta x^2 + \Delta y^2}} = \lim_{(\Delta x,\Delta y)\to \mathbf{0}} \frac{\Delta x^2}{\sqrt{\Delta x^2 + \Delta y^2}} = 0,$$

because

$$\frac{\Delta x^2}{\sqrt{\Delta x^2 + \Delta y^2}} \le \frac{\Delta x^2}{|\Delta x|} = |\Delta x|.$$

Example 4.3.3 highlights one problem for those readers who are new to the definition of vector-valued function derivatives. The notation Df(a,b)(x,y) looks very

complicated because it contains so many variables. One needs to look beyond the notation. We are simply defining a linear map T(x,y), but one that varies as we move from one point (a,b) to another. If the reader will bear with us, we shall soon introduce a more common and more compact notation for the derivative, namely the Jacobian matrix of partial derivatives, that will eliminate the excess of symbols. The advantage of the current notation however is that there is absolutely no ambiguity about it whereas the same cannot be said about some instances of the partial derivative notation.

Now, since we clearly do not want to compute derivatives via limits, we need some basic facts that will aid in their computation.

**4.3.4. Proposition**

(1) If f is a constant map, then Df = $\mathbf{0}$.
(2) If f is a linear transformations, then Df = f.
(3) The derivative at the point (a,b) of the map f(x,y) = xy is defined by Df(a,b)(x,y) = bx + ay.
(4) If f and g are two differentiable functions with the same domain and range, then f + g is differentiable and D(f + g) = Df + Dg.

**Proof.** The proposition follows in a straightforward manner from the definition of the derivative and is left as an exercise (Exercise 4.3.1).

The next proposition reduces the problem of finding the derivative of a vector-valued function to finding the derivative of real-valued functions of several variables.

**4.3.5. Proposition.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ and let $f:\mathbf{U} \to \mathbf{R}^m$. If $f(\mathbf{p}) = (f_1(\mathbf{p}), \ldots, f_m(\mathbf{p}))$, where $f_i:\mathbf{R}^n \to \mathbf{R}$, then f is differentiable at $\mathbf{p}$ if and only if each $f_i$ is and

$$Df(\mathbf{p})(\mathbf{h}) = (Df_1(\mathbf{p})(\mathbf{h}), \ldots, Df_m(\mathbf{p})(\mathbf{h})).$$

**Proof.** This is again straightforward. See [Spiv65].

**4.3.6. Theorem.** (The Chain Rule) Let $\mathbf{U}$ and $\mathbf{V}$ be open sets in $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively. Suppose that we have maps $f:\mathbf{U} \to \mathbf{R}^m$, $g:\mathbf{V} \to \mathbf{R}^k$ and that $f(\mathbf{U}) \subseteq \mathbf{V}$. If f is differentiable at $\mathbf{p}$ in $\mathbf{U}$ and g is differentiable at $\mathbf{q} = f(\mathbf{p})$, then the composite $g \circ f:\mathbf{R}^n \to \mathbf{R}^k$ is differentiable at $\mathbf{p}$ and

$$D(g \circ f)(\mathbf{p}) = Dg(\mathbf{q}) \circ Df(\mathbf{p}).$$

**Proof.** See [Spiv65]. The proof is similar to the one for functions of one variable.

It is very important that one understands the chain rule for derivatives. Not only does it lead to many differentiation formulas (starting with only a few basic ones) but it gets used over and over again in practical problems. Lots of nice formulas would be impossible without it. One immediate corollary is that the derivative of vector-valued functions satisfies the usual properties of a derivative.

**4.3.7. Corollary.** Let $\mathbf{U}$ be an open set in $\mathbf{R}^n$ and let $f, g: \mathbf{U} \to \mathbf{R}$ be functions that are differentiable at $\mathbf{p} \in \mathbf{U}$.

(1) $D(f + g)(\mathbf{p}) = Df(\mathbf{p}) + Dg(\mathbf{p})$
(2) $D(fg)(\mathbf{p}) = f(\mathbf{p})Dg(\mathbf{p}) + g(\mathbf{p})Df(\mathbf{p})$
(3) If $g(\mathbf{p}) \neq 0$, then

$$D(f/g)(\mathbf{p}) = \frac{g(\mathbf{p})Df(\mathbf{p}) - f(\mathbf{p})Dg(\mathbf{p})}{g(\mathbf{p})^2}.$$

**Proof.** We prove (1) to show how the chain rule gets used. The rest are left as an exercise. Define functions $\mu(\mathbf{p}) = (f(\mathbf{p}),g(\mathbf{p}))$ and $\sigma(\mathbf{p},\mathbf{q}) = \mathbf{p} + \mathbf{q}$. Clearly, $(f + g)(\mathbf{p}) = \sigma(\mu(\mathbf{p}))$. Therefore, the chain rule, Proposition 4.3.4(2), and Proposition 4.3.5 and imply that

$$D(f+g)(\mathbf{p}) = D(\sigma \circ \mu)(\mathbf{p}) = D\sigma(\mu(\mathbf{p})) \circ D\mu(\mathbf{p}) = \sigma \circ (Df(\mathbf{p}), Dg(\mathbf{p})) = Df(\mathbf{p}) + Dg(\mathbf{p}).$$

**4.3.8. Example.** Let $f(u) = (\sin u, u^2)$, $g(x,y) = x^2 + 3xy$, and $G(u) = g(f(u))$. Find the derivative DG.

**Solution.** In this problem, it is of course easy to compose the functions f and g to get

$$G(u) = \sin^2 u + 3u^2 \sin u$$

and

$$G'(u) = 2\sin u \cos u + 3u^2 \cos u + 6u \sin u.$$

On the other hand, using the chain rule we do not need to compute G(u) directly. Note that

$$Df(a)(h) = ((\cos a)h, (2a)h)$$

and

$$Dg(c,d)(h,k) = 2ch + (3dh + 3ck).$$

It follows that

$$\begin{aligned} DG(a)(h) &= Dg(\sin a, a^2)((\cos a)h, (2a)h) \\ &= (2\sin a \cos a)h + (3a^2 \cos a)h + (6a \sin a)h, \end{aligned}$$

which agrees with the first answer.

Although the chain rule lets us compute the derivatives of many functions, its direct use is rather messy. To make computations still easier, we use the chain rule to determine the matrix for the linear transformation $Df(\mathbf{p})$.

Let $f: \mathbf{R}^n \to \mathbf{R}$ and let $\mathbf{p} \in \mathbf{R}^n$.

**Definition.** The limit

$$\lim_{h\to 0} \frac{f(\mathbf{p}+h\mathbf{e}_i)-f(\mathbf{p})}{h} = \lim_{h\to 0} \frac{f(p_1,\ldots,p_{i-1},p_i+h,p_{i+1},\ldots,p_n)-f(p_1,\ldots,p_n)}{h},$$

if it exists, is called the *ith partial derivative of f of order 1 at* $\boldsymbol{p}$ and denoted by $D_i f(\mathbf{p})$ or $\partial f/\partial x_i(\mathbf{p})$.

Note that the ith partial derivative is just the ordinary derivative $h'(0)$ of the composite function $h(t) = f(g(t))$, where $g(t) = \mathbf{p} + t\mathbf{e}_i$.

Since partial derivatives are again functions, one can take partial derivatives of those to get the higher partial derivatives.

**Definition.** If $k > 1$, then define the (*mixed*) *partial derivative* $D_{i_1,\ldots,i_k} f(\mathbf{p})$ *of order k* recursively by:

$$D_{i_1,\ldots,i_k} f(\mathbf{p}) = D_{i_k}(D_{i_1,\ldots,i_{k-1}} f)(\mathbf{p}).$$

Does it matter in which order the partial derivatives are taken? Usually not.

**4.3.9. Theorem.** If $D_{i,j}f$ and $D_{j,i}f$ are continuous in an open neighborhood containing $\mathbf{p}$, then

$$D_{i,j}f(\mathbf{p}) = D_{j,i}f(\mathbf{p}).$$

**Proof.** See [Spiv65].

Some common notation for partial derivatives is:

$\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$, and $\dfrac{\partial f}{\partial z}$ or $f_x, f_y$, and $f_z$ for $D_1 f, D_2 f$, and $D_3 f$, respectively.

If $f: \mathbf{R} \to \mathbf{R}$, then one does not usually write $\dfrac{\partial f}{\partial x}$ but reverts to $\dfrac{df}{dx}$ or $f'(x)$.

$\dfrac{\partial^2 f}{\partial x^2}, \dfrac{\partial^2 f}{\partial x \partial y}$, or $f_{xx}, f_{xy}, \ldots$ for $D_{1,1}f, D_{2,1}f, \ldots$

Finally, the following notation often comes in handy to simplify expressions.

**Notation.** Let $f: \mathbf{R}^n \to \mathbf{R}^m$ be differentiable. If $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, then

$$\frac{\partial f}{\partial x_i} \quad \text{will denote} \quad \left(\frac{\partial f_1}{\partial x_i}, \frac{\partial f_2}{\partial x_i}, \ldots, \frac{\partial f_m}{\partial x_i}\right).$$

**Definition.** Let $\mathbf{A}$ be an open set in $\mathbf{R}^n$ and let $f: \mathbf{A} \to \mathbf{R}$. If f is continuous, it is said to be of *class* $C^0$. Let $k \geq 1$. The function f is said to be of *class* $C^k$ if its partial deriv-

atives of order r exist **and** are continuous for $1 \le R \le k$. The set of such functions is denoted by $C^k(\mathbf{A})$. If $f: \mathbf{A} \to \mathbf{R}^m$, then f is said to be of *class* $C^k$ if all the component functions are of class $C^k$. $C^1$ functions are often called *continuously differentiable* functions. If f is of class $C^k$ for all k, then we say that f is of *class* $C^\infty$. A $C^\infty$ function is also called a *smooth* function.

**4.3.10. Theorem.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$. If $f: \mathbf{U} \to \mathbf{R}^m$ is differentiable at $\mathbf{p}$, then the jth partial derivatives $D_j f_i(\mathbf{p})$ exist for all i and the $m \times n$ matrix $(D_j f_i(\mathbf{p}))$ is the **transpose** of what we have called the matrix for the linear transformation $Df(\mathbf{p})$ with respect to the standard bases of $\mathbf{R}^n$ and $\mathbf{R}^m$.

**Proof.** This is an easy consequence of the definitions and the chain rule. See [Spiv65] for details.

**Definition.** With the notation and hypotheses of Theorem 4.3.10, the matrix $(D_j f_i(\mathbf{p}))$ is called the *Jacobian matrix* and will be denoted by $f'(\mathbf{p})$. Using the more common partial derivative notation,

$$f' = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

If $m = n$, then the determinant of the $n \times n$ Jacobian matrix $f'(\mathbf{p})$ is called the *Jacobian* of f at $\mathbf{p}$. If $m = 1$, then it is often convenient to treat the $1 \times n$ matrix $f'(\mathbf{p})$ as a vector in $\mathbf{R}^n$.

**Note.** With our definition, the matrix $f'(\mathbf{p})$, thought of as a transformation, will act on vectors on the **right** and not on the left as is the case everywhere else in this book. The reason for departing from our usual convention regarding the matrix of a linear transformation and using the transpose here is to be compatible with how most people define the Jacobian matrix in terms of partials. If $m = n = 1$, the new notation $f'(\mathbf{p})$ will also be compatible with the old notation for the derivative of a function of one variable if we think of a number as a $1 \times 1$ matrix.

**4.3.11. Theorem.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ and let $f: \mathbf{U} \to \mathbf{R}^m$ and assume that each $D_j f_i$ exists in a neighborhood of a point $\mathbf{p}$ and is continuous at $\mathbf{p}$. Then f is differentiable at $\mathbf{p}$.

**Proof.** See [Spiv65].

Assuming that one can take partial derivatives of ordinary real-valued functions, we now show via some examples how easy it is to find derivatives of vector-valued functions and how the derivatives really do correspond to approximating the func-

tions by linear functions. In the three examples below, the functions should be thought of intuitively as parameterizations of certain sets. The derivative will then correspond to the tangent line or plane at the appropriate points of those sets.

**4.3.12. Example.** To find the derivative of $p(\theta) = (\cos\theta, \sin\theta)$ at $\theta = \pi/4$.

**Solution.** By Theorem 4.3.10

$$p'(a) = \begin{pmatrix} -\sin a \\ \cos a \end{pmatrix}.$$

Therefore,

$$Dp(a)(h) = (-h\sin a, h\cos a) \quad \text{and} \quad Dp\left(\frac{\pi}{4}\right)(h) = \left(-\frac{1}{\sqrt{2}}h, \frac{1}{\sqrt{2}}h\right).$$

The graph of the function

$$T(u) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) + \left(-\frac{1}{\sqrt{2}}\left(u - \frac{\pi}{4}\right), \frac{1}{\sqrt{2}}\left(u - \frac{\pi}{4}\right)\right)$$

is in fact the tangent line to the circle at $p(\pi/4)$.

**4.3.13. Example.** To find the derivative of $f(x,y) = x^2 + y^2$ at $(x,y) = (1,0)$.

**Solution.** By Theorem 4.3.10, $f'(a,b) = (2a,2b)$. Therefore,

$$Df(a,b)(h,k) = 2ah + 2bk \quad \text{and} \quad Df(1,0)(h,k) = 2h.$$

The graph of

$$T(x,y) = f(1,0) + 2(x-1) = 2x - 1$$

is the tangent plane to the graph of f at (1,0).

**4.3.14. Example.** To find the derivative of $f(\theta,z) = (\cos\theta, \sin\theta, z)$ at (0,2).

**Solution.** By Theorem 4.3.10

$$f'(a,b) = \begin{pmatrix} -\sin a & 0 \\ \cos a & 0 \\ 0 & 1 \end{pmatrix},$$

so that $Df(a,b)(h,k) = ((-\sin a)h, (\cos a)h, k)$. Clearly, the map

$$T(\theta,z) = (1,0,z) + (0,\theta,z)$$

is the tangent plane at (1,0,2) to the cylinder of radius 1 centered on the z-axis, which is the surface parameterized by f.

Up to now we have been dealing with the derivative of a function as a linear map (which is what it is), but the reader probably has also noticed that this is a little cumbersome and we end up with complicated looking expressions even if we use the Jacobian matrix. This is not how most people deal with derivatives. To get more of the look and feel of how most people really work, we use the following consequence of the chain rule.

**4.3.15. Theorem.** Let $g_1, \ldots, g_m : \mathbf{R}^n \to \mathbf{R}$, $f : \mathbf{R}^m \to \mathbf{R}$ be functions that are continuously differentiable at $\mathbf{a} \in \mathbf{R}^n$ and $(g_1(\mathbf{a}), \ldots, g_m(\mathbf{a})) \in \mathbf{R}^m$, respectively. Define $F : \mathbf{R}^n \to \mathbf{R}$ by $F(\mathbf{x}) = f(g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}))$. Then

$$D_i F(\mathbf{a}) = \sum_{j=1}^{m} D_j f(g_1(\mathbf{a}), \ldots, g_m(\mathbf{a})) D_i g_j(\mathbf{a}). \tag{4.5}$$

**Proof.** See [Spiv65].

Equation (4.5), which is really just the chain rule, comes in very handy when computing derivatives and is usually written informally as

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{m} \frac{\partial f}{\partial y_j} \frac{\partial y_j}{\partial x_i}. \tag{4.6}$$

Nevertheless, one needs to be aware of the fact that the Df notation of a derivative in equation (4.5) is more precise and one should always return to that if one has any problems carrying out a computation. Specifically, although notation of the type $\partial f/\partial x$ is the more common notation for the ith partial derivative it can sometimes be ambiguous whereas the other notation $D_i f(\mathbf{p})$ is **not**. A typical case where ambiguities can arise is in the application of the chain rule like in equation (4.6). For example, if f(u,v) is a function and u = g(x,y) and v = h(x,y), then equation (4.6) turns into

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x},$$

Although this way of writing the chain rule makes it easy to remember the rule, the formula must be interpreted carefully. For one thing, the f's appearing on the right hand side of the equation are different from the f on the left hand side. In the first case, we are considering f to be a function of u and v and, in the second, we are considering f to be a function of x and y. One should really write

$$\frac{\partial F}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x},$$

where F(x,y) = f(u(x,y),v(x,y)). Sloppiness may be acceptable, but only as long as one knows how to express things correctly when needed.

**4.3.16. Example.** Given $z = x^2 + 3xy$, $x = \sin u$, and $y = u^2$, to find dz/du. Note that this is really just a restatement of the problem in Example 4.3.8.

**Solution.** We have

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{dy}{du} = (2x+3y)\cos u + (3x)2u = 2\sin u\cos u + 3u^2\cos u + 6\sin u.$$

This answer clearly agrees with the earlier one.

**4.3.17. Example.** Let u = f(x − ct) + g(x + ct). To show that

$$c^2\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}.$$

**Solution.** Define r(x,t) = x − ct and s(x,t) = x + ct. Then u(x,t) = f(r(x,t)) + g(s(x,t)). The result follows from the following computations:

$$\frac{\partial u}{\partial x}(x,t) = f'(r(x,t))\frac{\partial r}{\partial x}(x,t) + g'(r(x,t))\frac{\partial s}{\partial x}(x,t) = f'(r(x,t)) + g'(r(x,t))$$
$$\frac{\partial u}{\partial t}(x,t) = f'(r(x,t))\frac{\partial r}{\partial t}(x,t) + g'(r(x,t))\frac{\partial s}{\partial t}(x,t) = -cf'(r(x,t)) + cg'(r(x,t))$$

and

$$\frac{\partial^2 u}{\partial x^2}(x,t) = f''(r(x,t))\frac{\partial r}{\partial x}(x,t) + g''(r(x,t))\frac{\partial s}{\partial x}(x,t) = f''(r(x,t)) + g''(r(x,t))$$
$$\frac{\partial^2 u}{\partial t^2}(x,t) = -cf''(r(x,t))\frac{\partial r}{\partial t}(x,t) + cg''(r(x,t))\frac{\partial s}{\partial t}(x,t) = c^2f''(r(x,t)) + c^2g''(r(x,t))$$

**Definition.** Let $f: \mathbf{R}^n \to \mathbf{R}$ be a function that is differentiable at a point $\mathbf{p}$. Then the *gradient of f at* $\boldsymbol{p}$, denoted by $\nabla f(\mathbf{p})$, is defined to be the Jacobian matrix $f'(\mathbf{p})$, that is,

$$\nabla f(\mathbf{p}) = (D_1 f(\mathbf{p}), \ldots, D_n f(\mathbf{p})).$$

A generalization of partial derivatives is the directional derivative.

**Definition.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ and let $f: \mathbf{U} \to \mathbf{R}$. Let $\mathbf{v}$ be an arbitrary nonzero vector. If $\mathbf{p} \in \mathbf{U}$ and if the limit

$$\lim_{h\to 0}\frac{f(\mathbf{p}+h\mathbf{v}) - f(\mathbf{p})}{h}$$

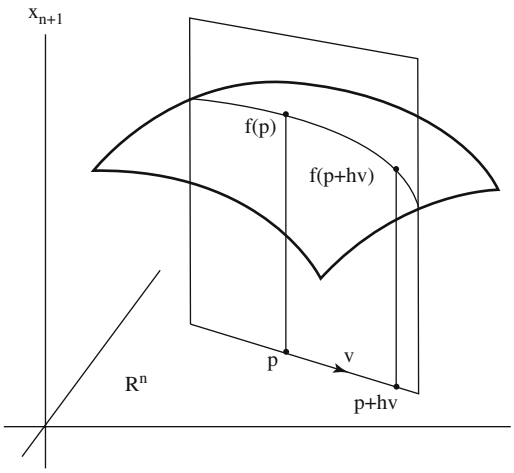



**Figure 4.9.** The directional derivative.

exists, then that limit is called the *directional derivative of f in the direction* $\boldsymbol{v}$ and is denoted by $D_{\mathbf{v}}f(\mathbf{p})$.

The directional derivative at a point $\mathbf{p}$ is essentially the derivative of the curve obtained by intersecting the graph of f with a vertical plane through $\mathbf{p}$. See Figure 4.9. Some books require that $\mathbf{v}$ be a unit vector. Although this is a natural requirement in some application (see Examples 4.3.19 and 4.3.20 below), there is no reason to assume this in the definition. In fact, part (2) of the next proposition shows a useful linear relationship in the general case.

**4.3.18. Proposition**

(1) $D_{\mathbf{v}}f(\mathbf{p}) = \nabla f(\mathbf{p}) \bullet \mathbf{v}$. In particular, $D_{\mathbf{e}_i} f(\mathbf{p}) = D_i f(\mathbf{p}) = \dfrac{\partial f}{\partial x_i}(\mathbf{p})$.

(2) The directional derivative is a linear function of its direction vector, that is, if $\mathbf{v}, \mathbf{w} \in \mathbf{R}^n$ and $a, b \in \mathbf{R}$, then

$$D_{a\mathbf{v}+b\mathbf{w}}f = aD_{\mathbf{v}}f + bD_{\mathbf{w}}f.$$

**Proof.** To prove part (1), define functions $\tau$ and g by

$$\tau(h) = \mathbf{p} + h\mathbf{v} \quad \text{and} \quad g(h) = f(\tau(h)) = f(\mathbf{p} + h\mathbf{v}).$$

The chain rule implies that

$$g'(h) = \nabla f(\tau(h)) \bullet \tau'(h) = \nabla f(\tau(h)) \bullet \mathbf{v}.$$

But $D_{\mathbf{v}}f(p)$ is just $g'(0)$ and so we are done. Part (2) follows easily from (1) because

$$D_{a\mathbf{v}+b\mathbf{w}}f = \nabla f \bullet (a\mathbf{v} + b\mathbf{w}) = a\nabla f \bullet \mathbf{v} + b\nabla f \bullet \mathbf{w} = aD_{\mathbf{v}}f + bD_{\mathbf{w}}f.$$

**4.3.19. Example.** To find $D_{\mathbf{v}}f$ when $\mathbf{v} = (-1,2)$ and $f(x,y) = x^2y + e^{xy}$.

**Solution.** Since $\nabla f\,(x,y) = (2xy + ye^{xy}, x^2 + xe^{xy})$, it follows that

$$D_{(-1,2)}f(x,y) = \left(2xy + ye^{xy}, x^2 + xe^{xy}\right) \bullet (-1,2) = -2xy - ye^{xy} + 2x^2 + 2xe^{xy}.$$

**4.3.20. Example.** To find the directional derivative for

$$F(x,y) = e^x \cos x\, y$$

at (0,0) in the direction making an angle of 60 degrees with the x-axis.

**Solution.** The unit direction we want is $\mathbf{v} = (1/2)(1,\sqrt{3})$. Since $\nabla F(x,y) = (e^x \cos y, -e^x \sin y)$, it follows that

$$D_{\mathbf{v}}F(0,0) = \nabla F(0,0) \bullet \mathbf{v} = (1,0) \bullet \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \frac{1}{2}.$$

**4.3.21. Example.** To find the directional derivative of

$$F(x,y,z) = x^2yz^3$$

along the curve $\gamma(u) = (e^{-u}, 2\sin u + 1, u - \cos u)$ at the point $\gamma(0)$ on the curve.

**Solution.** What we are after is the directional derivative of F in the direction of the unit tangent vector to the curve $\gamma(u)$ at 0. We shall see in Section 8.4 that the tangent vector $\mathbf{v}(u)$ to the curve at u can be obtained by differentiating the component functions of the curve, so that $\mathbf{v}(u) = (-e^u, 2\cos u, 1+\sin u)$ and $\mathbf{v}(0) = (-1,2,1)$. Let $\mathbf{u}$ be the unit vector in the direction $\mathbf{v}(0)$. Since $\nabla F(x,y,z) = (2xyz^3, x^2z^3, 3x^2yz^2)$ and $\gamma(0) = (1,1,-1)$, our answer is

$$D_{\mathbf{u}}F(1,1,-1) = \nabla F(1,1,-1) \bullet \mathbf{u} = (-2,-1,3) \bullet \frac{1}{\sqrt{6}}(-1,2,1) = \sqrt{\frac{3}{2}}.$$

Here are two more basic theorems for vector-valued functions that extend well-known results from the case of ordinary functions of one variable.

**4.3.22. Theorem.** (The Generalized Mean Value Theorem) Let $f: \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable function. If $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$, then

$$f(\mathbf{q}) - f(\mathbf{p}) = Df(\mathbf{p}^*)(\mathbf{q} - \mathbf{p})$$

for some $\mathbf{p}^* \in [\mathbf{p},\mathbf{q}]$.

**Proof.** See [Buck78].

The next theorem is a generalization of Taylor's theorem. We shall only state it for the 2-variable case. First, it is convenient to define a differential operator

$$a\frac{\partial}{\partial x}+b\frac{\partial}{\partial y}$$

which, when applied to a real-valued function f(x,y), gives

$$a\frac{\partial f}{\partial x}(x,y)+b\frac{\partial f}{\partial y}(x,y).$$

For example,

$$\left(a\frac{\partial}{\partial x}+b\frac{\partial}{\partial y}\right)^2(f)(x,y)=a^2\frac{\partial^2 f}{\partial x^2}(x,y)+2ab\frac{\partial^2 f}{\partial x\partial y}(x,y)+b^2\frac{\partial^2 f}{\partial y^2}(x,y).$$

**Definition.** Let f(x,y) be of class $C^n$ in a neighborhood of a point $(x_0,y_0)$. Then

$$g(x,y)=f(x_0,y_0)+\sum_{k=1}^{n}\frac{1}{n!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^k(f)(x,y)$$

is called the *Taylor polynomial of f of degree n at $x_0$*.

**4.3.23. Example.** To find the Taylor polynomial g(x,y) of degree 2 at (1,2) for the function

$$f(x,y)=2x^3+5xy^2.$$

**Solution.** Now

$$\frac{\partial f}{\partial x}=6x^2,\quad \frac{\partial f}{\partial y}=10xy,\quad \frac{\partial^2 f}{\partial x^2}=12x,\quad \frac{\partial^2 f}{\partial x\partial y}=0,\quad \text{and}\quad \frac{\partial^2 f}{\partial y^2}=10x,$$

so that

$$g(x,y)=18+6(x-1)+20(y-2)+\frac{1}{2}\left[12(x-1)^2+2\cdot 0\cdot(x-1)(y-2)+10(y-2)^2\right].$$

**4.3.24. Theorem.** (The Taylor Polynomial Theorem) Let f(x,y) be of class $C^{n+1}$ in a neighborhood of a point $(x_0,y_0)$. Then

$$f(x,y)=f(x_0,y_0)+\sum_{k=1}^{n}\frac{1}{n!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^k(f)(x,y)+R_{n+1},$$

where

$$R_{n+1}=\frac{1}{(n+1)!}\left[(x-x_0)\frac{\partial}{\partial x}+(y-y_0)\frac{\partial}{\partial y}\right]^{n+1}(f)(x^*,y^*)$$

for some point $(x^*,y^*)$ on the line segment from $(x_0,y_0)$ to $(x,y)$.

**Proof.** Let $\mathbf{p} = (x,y)$, $\mathbf{p}_0 = (x_0,y_0)$, and define

$$F(t) = f(\mathbf{p}_0 + t(\mathbf{p} - \mathbf{p}_0)),\ t \in [0,1].$$

The theorem follows easily from the chain rule and the basic Taylor polynomial for functions of one variable applied to F (Theorem D.2.3). See [Buck78] .

So far in this section differentiability was a notion that was defined only for a function f whose domain **A** was an open set; however, one can define the derivative of a function also in cases where its domain is a more general set. Basically, all one has to be able to do is extend the function f to a function F defined on an open set containing **A**. One then defines the derivative of f to be the derivative of F and shows that this value does not depend on the extension F one has chosen. In fact, one only really needs local extensions, that is, for every point in **A** we need to be able to extend f to a differentiable function on a neighborhood of that point.

**Definition 1.** Let **A** be an arbitrary subset of $\mathbf{R}^n$. A map $f: \mathbf{A} \to \mathbf{R}^m$ is said to be *of class $C^k$* or a *$C^k$ map on **A*** if there exists an open neighborhood **U** of **A** in $\mathbf{R}^n$ and a $C^k$ map

$$F:\ \mathbf{U} \to \mathbf{R}^m$$

that extends f, that is, $f = F \mid (\mathbf{U} \cap \mathbf{A})$. If $k \geq 1$, then the *rank of f at a point* **p** is the rank of DF at **p**.

**Definition 2.** Let **A** be an arbitrary subset of $\mathbf{R}^n$. A map $f: \mathbf{A} \to \mathbf{R}^m$ is said to be *of class $C^k$* or a *$C^k$ map at a point* **p** in **A** if there exists a neighborhood $\mathbf{U_p}$ of **p** in $\mathbf{R}^n$ and a $C^k$ map

$$F_{\mathbf{p}}:\ \mathbf{U_p} \to \mathbf{R}^m$$

that extends $f \mid (\mathbf{U_p} \cap \mathbf{A})$. If $k \geq 1$, then the *rank of f at the point* **p** is the rank of DF at **p**. The map f is a *$C^k$ map* if it is of class $C^k$ at every point **p** in **A**.

**4.3.25. Theorem.** The definitions of $C^k$ maps on a set or at a point are well defined. The two definitions of $C^k$ maps on a set are equivalent. The notion of rank is well defined in all cases. If the set **A** is open, then the definitions agree with the earlier definition of differentiability and rank.

**Proof.** For details see [Munk61].

Notice that neither definition actually defined a derivative although we did define the rank of the map. Since the extensions F are not unique, it is not possible to define a derivative in general. In certain common cases, such as rectangles or disks where boundary points have nice "half-space" neighborhoods, the derivative **is** defined uniquely. Actually, in such cases, one could simply define the derivative at such a point

as "one-sided" limits, extending the idea of the derivative at the endpoints of a function defined on a closed interval [a,b]. All the theorems and definitions in this section will be applicable.

## 4.4 The Inverse and Implicit Function Theorem

**Definition.** Let **U** and **V** be open subsets of $\mathbf{R}^n$. A $C^k$ map $f: \mathbf{U} \to \mathbf{V}$, $k = \infty$ or $k \geq 1$, that has a $C^k$ inverse is called a $C^k$ *diffeomorphism* of **U** onto **V**. A $C^\infty$ diffeomorphism will be called simply a *diffeomorphism*.

Because diffeomorphisms are one-to-one and onto maps, one can think of them as defining a change of coordinates. Another definition that often comes in handy is the following:

**Definition.** Let **U** be an open subset of $\mathbf{R}^n$ and let $f: \mathbf{U} \to \mathbf{R}^n$. If $\mathbf{p} \in \mathbf{U}$, then f is called a *local ($C^k$) diffeomorphism at* $\boldsymbol{p}$ if f is a ($C^k$) diffeomorphism of an open neighborhood of **p** onto an open neighborhood of f(**p**).

**4.4.1. Lemma.** If a differentiable map $f: \mathbf{B}^n(r) \to \mathbf{R}^n$ satisfies

$$\left|\frac{\partial f_i}{\partial x_j}\right| \leq b, \quad \text{for all i and j,}$$

then it satisfies the Lipschitz condition

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq bn|\mathbf{x} - \mathbf{y}|, \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbf{B}^n.$$

**Proof.** Use Taylor series and the mean-value theorem.

**4.4.2. Theorem.** (The Inverse Function Theorem) Let **U** be an open subset of $\mathbf{R}^n$. Let $f: \mathbf{U} \to \mathbf{R}^n$ be a $C^k$ function, $k \geq 1$, and assume that $Df(\mathbf{x}_0)$ is nonsingular at $\mathbf{x}_0 \in \mathbf{U}$. Then f is a local $C^k$ diffeomorphism.

**Outline of proof.** By composing f with linear maps if necessary one may assume that $\mathbf{x}_0 = f(\mathbf{x}_0) = \mathbf{0}$ and that $Df(\mathbf{x}_0)$ is the identity map. Next, let $g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{x}$. It follows that $Dg(\mathbf{0})$ is the zero map and since f is at least $C^1$, there is a small neighborhood $\mathbf{B}^n(r)$ about the origin so that

$$\left|\frac{\partial g_i}{\partial x_j}\right| \leq \frac{1}{2n}.$$

**Claim.** For each $\mathbf{y} \in \mathbf{B}^n(r/2)$ there is a unique $\mathbf{x} \in \mathbf{B}^n(r)$ such that $f(\mathbf{x}) = \mathbf{y}$.

See Figure 4.10. To prove the existence of **x** note that Lemma 4.4.1 implies that

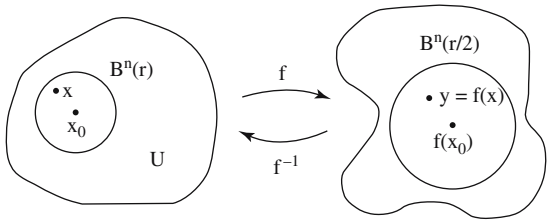



**Figure 4.10.** Proving the inverse function theorem.

$$|g(\mathbf{x}) - g(\mathbf{x}_0)| \le \frac{1}{2}|\mathbf{x} - \mathbf{x}_0| \tag{4.7}$$

for $\mathbf{x} \in \mathbf{B}^n(r)$. Define $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{x}_1 = \mathbf{y}$, and $x_{m+1} = \mathbf{y} - g(\mathbf{x}_m)$, for $m \ge 1$. Our hypotheses imply that

$$|\mathbf{x}_m - \mathbf{x}_{m-1}| \le \frac{1}{2}|\mathbf{x}_{m-2} - \mathbf{x}_{m-1}|,$$

and so $|\mathbf{x}_m| \le 2|\mathbf{y}|$ for all k. It follows that the $\mathbf{x}_m$ converge to a point $\mathbf{x}$ with $|\mathbf{x}| \le 2|\mathbf{y}|$, that is, $\mathbf{x} \in \mathbf{B}^n(r)$. Furthermore, $\mathbf{x} = \mathbf{y} - g(\mathbf{x})$, so that $f(\mathbf{x}) = \mathbf{y}$. To prove that $\mathbf{x}$ is unique, assume that $f(\mathbf{x}_1) = \mathbf{y}$. Then

$$\mathbf{x} - \mathbf{x}_1 = g(\mathbf{x}_1) - g(\mathbf{x}) \le \frac{1}{2}|\mathbf{x} - \mathbf{x}_1|,$$

so that $\mathbf{x} - \mathbf{x}_1 = 0$. The claim is proved.

The claim shows that

$$f^{-1}\colon \mathbf{B}^n\left(\frac{r}{2}\right) \to \mathbf{B}^n(r)$$

exists. The map $f^{-1}$ is continuous because

$$|f(\mathbf{x}) - f(\mathbf{x}_1)| \ge |\mathbf{x} - \mathbf{x}_1| - |g(\mathbf{x}) - g(\mathbf{x}_1)| \ge \frac{1}{2}|\mathbf{x} - \mathbf{x}_1|$$

implies that

$$|\mathbf{y} - \mathbf{y}_1| \ge \left|f^{-1}(\mathbf{y}) - f^{-1}(\mathbf{y}_1)\right|.$$

We still need to show that $f^{-1}$ is differentiable in addition to being continuous.

Since f is differentiable, we have that

$$f(\mathbf{x}) = f(\mathbf{x}_1) + Df(\mathbf{x}_1)(\mathbf{x} - \mathbf{x}_1) + h(\mathbf{x}, \mathbf{x}_1), \tag{4.8}$$

where

$$\frac{h(\mathbf{x},\mathbf{x}_1)}{|\mathbf{x}-\mathbf{x}_1|} \to 0 \quad \text{as} \quad \mathbf{x} \to \mathbf{x}_1.$$

We must show that an equation similar to (4.8) holds for $f^{-1}$. The obvious candidate for $Df^{-1}$ is $A = (Df)^{-1}$. Applying A to both sides of equation (4.8) gives

$$A(\mathbf{y}-\mathbf{y}_1) + A(h_1(\mathbf{y},\mathbf{y}_1)) = f^{-1}(\mathbf{y}) - f^{-1}(\mathbf{y}_1),$$

where $h_1(\mathbf{y},\mathbf{y}_1) = -h(f^{-1}(\mathbf{y}),f^{-1}(\mathbf{y}_1))$. Now

$$\frac{h_1(\mathbf{y},\mathbf{y}_1)}{|\mathbf{y}-\mathbf{y}_1|} = -\frac{h_1(\mathbf{y},\mathbf{y}_1)}{|\mathbf{x}-\mathbf{x}_1|}\frac{|\mathbf{x}-\mathbf{x}_1|}{|\mathbf{y}-\mathbf{y}_1|}$$

and the right-hand side of this equation goes to zero (the first term goes to zero and the second is bounded by 2). This shows that $f^{-1}$ is differentiable. The continuity of $D(f^{-1})$ follows from the fact that the matrix for this map is defined by the composite of the maps

$$\mathbf{B}^n\left(\frac{r}{2}\right) \xrightarrow[f^{-1}]{} \mathbf{B}^n(r) \xrightarrow[\text{matrix for Df}]{} \mathbf{GL}(n,\mathbf{R}) \xrightarrow[\text{matrix inversion}]{} \mathbf{GL}(n,\mathbf{R}),$$

where we identify the space $\mathbf{GL}(n,\mathbf{R})$ of nonsingular real $n \times n$ matrices with $\mathbf{R}^{n^2}$. If f is of class $C^k$, then one can show in a similar fashion that $f^{-1}$ is also.

To prove the first application of the inverse function theorem we need two lemmas.

**4.4.3. Lemma.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ that contains the origin. Let $f:\mathbf{U} \to \mathbf{R}^m$, $n \le m$, be a $C^k$ map with $f(\mathbf{0}) = \mathbf{0}$ and $k \ge 1$. Assume that $Df(\mathbf{0})$ has rank n. Then there is a $C^k$ diffeomorphism g of one neighborhood of the origin in $\mathbf{R}^m$ onto another with $g(\mathbf{0}) = \mathbf{0}$ and such that

$$g(f(x_1,\ldots,x_n)) = (x_1,\ldots,x_n,0,\ldots,0)$$

holds in some neighborhood of the origin in $\mathbf{R}^n$.

**Proof.** The hypothesis that $Df(\mathbf{0})$ has rank n means that the $n \times m$ Jacobian matrix $(\partial f_i/\partial x_j)$ has rank n. Because we can interchange coordinates if necessary, there is no loss in generality if we assume that

$$\operatorname{rank}\left(\frac{\partial f_i}{\partial x_j}\right)_{1\le i,j\le n} = n.$$

Consider the map $F:\mathbf{U} \times \mathbf{R}^{m-n} \to \mathbf{R}^m$ defined by

$$F(x_1,\ldots,x_m) = f(x_1,\ldots,x_n) + (0,\ldots,0,x_{n+1},\ldots,x_m).$$

Since $F(x_1,\ldots,x_n,0,\ldots,0) = f(x_1,\ldots,x_n)$, F is an extension of f. Furthermore, the determinant of $(\partial F_i/\partial x_j)$ is just the determinant of

$$\left(\frac{\partial f_i}{\partial x_j}\right)_{1\le i,j\le n},$$

which is nonzero. The inverse function theorem now implies that F has a local inverse g that is a $C^k$ diffeomorphism of one neighborhood of the origin in $\mathbf{R}^m$ onto another. Therefore,

$$\begin{aligned} g(f(x_1,\ldots,x_n)) &= g(F(x_1,\ldots,x_n,0,\ldots,0)) \\ &= (x_1,\ldots,x_n), \end{aligned}$$

and the Lemma is proved.

**4.4.4. Lemma.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ that contains the origin. Let $f:\mathbf{U}\to\mathbf{R}^m$, $n \ge m$, be a $C^k$ map with $f(\mathbf{0}) = \mathbf{0}$ and $k \ge 1$. Assume that $Df(\mathbf{0})$ has rank m. Then there is a $C^k$ diffeomorphism h of one neighborhood of the origin in $\mathbf{R}^n$ onto another with $h(\mathbf{0}) = \mathbf{0}$ and such that

$$f(h(x_1,\ldots,x_n)) = (x_1,\ldots,x_m)$$

holds in some neighborhood of the origin in $\mathbf{R}^n$.

**Proof.** Again, by interchanging coordinates if necessary, we may assume that

$$\operatorname{rank}\left(\frac{\partial f_i}{\partial x_j}\right)_{1\le i,j\le m} = m.$$

Define $F:\mathbf{U}\to\mathbf{R}^n$ by

$$F(x_1,\ldots,x_n) = (f_1(x),\ldots,f_m(x),x_{m+1},\ldots x_n).$$

Our hypothesis implies that F has a nonsingular Jacobian matrix at the origin and hence a local inverse h. Let g be the natural projection of $\mathbf{R}^n$ onto $\mathbf{R}^m$, which sends $(x_1,\ldots,x_n)$ onto $(x_1,\ldots,x_m)$. Then $f = gF$ and

$$\begin{aligned} f(h(x_1,\ldots,x_n)) &= g(F(h(x_1,\ldots,x_n))) \\ &= g(x_1,\ldots,x_n) \\ &= (x_1,\ldots,x_m). \end{aligned}$$

This proves the Lemma.

The next theorem can be interpreted as saying that up to change of curvilinear coordinates maps $f:\mathbf{R}^n\to\mathbf{R}^m$ basically look like the natural projection $(x_1,\ldots,x_n)\to(x_1,\ldots,x_k,0,\ldots,0)$ for appropriate k. Compare this result with Theorem 1.11.7, which deals with linear maps.

**4.4.5. Theorem.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$ that contains the origin. Let $f:\mathbf{U}\to\mathbf{R}^m$ be a $C^s$ map with $f(\mathbf{0}) = \mathbf{0}$ and $s \ge 1$. Assume that $Df(\mathbf{x})$ has rank k for all $\mathbf{x}$ in $\mathbf{U}$.

Then there exist $C^s$ diffeomorphisms h and g of neighborhoods of the origins in $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, such that

$$g(f(h(x_1,\ldots,x_n))) = (x_1,\ldots,x_k,0,\ldots,0)$$

in some neighborhood of the origin in $\mathbf{R}^n$.

**Proof.** The proof divides into two cases.

**Case 1.** $n \geq m$.

Let $\pi:\mathbf{R}^m \to \mathbf{R}^k$ be the natural projection. We may assume without loss of generality that the matrix

$$\left(\frac{\partial f_i}{\partial x_j}\right)_{1\leq i,j\leq k}$$

is nonsingular on $\mathbf{U}$. It follows from Lemma 4.4.4 (applied to $\pi$ of) that there is a $C^s$ diffeomorphism h such that

$$f(h(x_1,\ldots,x_n) = (x_1,\ldots,x_k,\phi_1(x_1,\ldots,x_n),\ldots,\phi_{m-k}(x_1,\ldots,x_n)).$$

Since the rank of $D(f\circ h)$ is k, we must have $\partial\phi_i/\partial x_j = 0$ for $j > k$. It follows that the $\phi_i$ are independent of $x_{k+1}, \ldots, x_n$. Now define a map $f_1:\mathbf{R}^k \to \mathbf{R}^m$ by

$$f_1(x_1,\ldots,x_k) = (x_1,\ldots,x_k,\phi_1(x_1,\ldots,x_k,0,\ldots,0),\ldots,\phi_{m-k}(x_1,\ldots,x_k,0,\ldots,0))$$

and apply Lemma 4.4.3.

**Case 2.** $n \leq m$.

This is proved similarly to Case 1 and is left as an exercise.

One aspect worth noting about the hypotheses of Theorem 4.4.5 is that it is not enough to simply assume that the map f has rank k at the origin. One needs to know that this holds in a neighborhood of the origin. In the special case where Df has maximal rank, then the assumption of rank k at the origin **is** enough because this by itself implies that Df will have maximal rank in a neighborhood. This explains the slight difference in hypotheses between Lemmas 4.4.3 and 4.4.4 and Theorem 4.4.5. It is worth summarizing an aspect of these observations.

**4.4.6. Theorem.** Let $\mathbf{U}$ be an open subset of $\mathbf{R}^n$. Let $f:\mathbf{U} \to \mathbf{R}^m$, $n \leq m$, be a $C^k$ map, $k \geq 1$. Let $\mathbf{p}$ be a point of $\mathbf{U}$ and assume that $Df(\mathbf{p})$ has rank n. Then there is a neighborhood of $\mathbf{V}$ of $\mathbf{p}$ in $\mathbf{U}$ so that $f|\mathbf{V}$ is one-to-one.

Implicit function theorems are another major application of the inverse function theorem. They have to do with solving equations of the form

$$f(x,y) = 0 \tag{4.9}$$

for x or y in terms of the other variable. We can think of equation (4.9) defining either x or y in terms of the other implicitly. The obvious question is under what conditions we can in fact think of x as a function of y or, conversely, y a function of x. A good example is

$$f(x,y) = x^2 + y^2 - 1 \tag{4.10}$$

in which case equation (4.9) defines the unit circle. A neighborhood of the point **A** = (0,1) on the circle is clearly the graph of the function

$$y(x) = \sqrt{1 - x^2}. \tag{4.11}$$

The variable x is **not** a function of y in any such neighborhood because functions are single-valued. On the other hand, a neighborhood of the point **B** = (1,0) on the circle is clearly the graph of the function

$$x(y) = \sqrt{1 - y^2}. \tag{4.12}$$

Here, the variable y is **not** a function of x in any neighborhood. The point **C** = $(1/\sqrt{2}, 1/\sqrt{2})$ is much nicer because we can solve for either x **or** y. See Figure 4.11. The difference between these points is that the tangent line at **A** and **B** is horizontal and vertical, respectively. The tangent line at **C** is neither. In fact, near **C** each point corresponds to a **unique** x and y value and the functions y(x) and x(y) given by equations (4.11) and (4.12), respectively, which are the local solutions to (4.9), are inverses of each other.

So what is a possible criterion that will guarantee that one can solve for one of the variables in equation (4.9)? Well, it is at points like **C** that have nonvertical tangents that we can guarantee both solutions. It is there that we can guarantee a unique x and y value for nearby points. For the points **A** and **B**, which have horizontal or vertical tangents (equivalently, derivatives of functions vanish or do not exist), we can at most guarantee one solution. Our theorem will only give us sufficient but not necessary conditions and not much can be said in general at points where appropriate derivatives vanish. They would have to be analyzed in special ways. Note that horizontal tangents are not necessarily bad because the curve

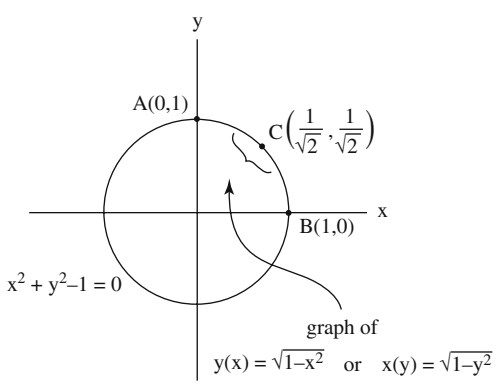



**Figure 4.11.** Solving for implicitly defined functions.

$$y - x^3 = 0$$

has a horizontal tangent at (0,0) but still can also be solved for both x and y in a neighborhood of that point.

**4.4.7. Theorem.** (The Implicit Function Theorem) Let

$$f\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$$

be a continuously differentiable function in an open set about a point $(\mathbf{a},\mathbf{b})$ and assume that

$$f(\mathbf{a},\mathbf{b}) = \mathbf{0}.$$

If the $m \times m$ matrix

$$M = (D_{n+i} f_j(\mathbf{a},\mathbf{b}))_{1 \le i,j \le m}$$

is nonsingular, then there exists an open neighborhood $\mathbf{A}$ about $\mathbf{a}$ in $\mathbf{R}^n$, an open neighborhood $\mathbf{B}$ about $\mathbf{b}$ in $\mathbf{R}^m$, and a differentiable function

$$g\colon \mathbf{A} \to \mathbf{B}$$

with the property that

$$f(\mathbf{x}, g(\mathbf{x})) = \mathbf{0}$$

for all $\mathbf{x}$ in $\mathbf{A}$.

**Proof.** Define a function

$$F\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$$

by

$$F(\mathbf{x},\mathbf{y}) = (\mathbf{x}, f(\mathbf{x},\mathbf{y})).$$

Our hypotheses imply that F is a continuously differentiable function whose derivative DF is nonsingular at $(\mathbf{a},\mathbf{b})$ because det $DF(\mathbf{a},\mathbf{b})$ = det M. The inverse function theorem (Theorem 4.4.2) now implies that F has an inverse G in an open neighborhood of $(\mathbf{a},\mathbf{0}) = F(\mathbf{a},\mathbf{b})$, which we may assume to have the form $\mathbf{A} \times \mathbf{B}$. It is easy to check that G has the form $G(\mathbf{x},\mathbf{y}) = (\mathbf{x}, k(\mathbf{x},\mathbf{y}))$ for some differentiable function k. Let

$$\pi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$$

be the natural projection. Then $\pi \circ F = f$ and

$$\begin{aligned} f(x,k(\mathbf{x},\mathbf{y})) &= (f \circ G)(\mathbf{x},\mathbf{y}) \\ &= ((\pi \circ F) \circ G)(\mathbf{x},\mathbf{y}) \\ &= (\pi \circ (F \circ G))(\mathbf{x},\mathbf{y}) \\ &= \pi(\mathbf{x},\mathbf{y}) \\ &= \mathbf{y}. \end{aligned}$$

Letting $g(\mathbf{x}) = k(\mathbf{x},\mathbf{0})$ proves the theorem.

Note that the chain rule for differentiation could easily be used to compute the derivative of the function g in Theorem 4.4.7.

Let us see how the implicit function theorem applies to our earlier example dealing with the unit circle and the function f in equation (4.10). We have that

$$\frac{\partial f}{\partial x} = 2x \quad \text{and} \quad \frac{\partial f}{\partial y} = 2y.$$

Evaluating $\partial f/\partial x$ and $\partial f/\partial y$ at the points **A**, **B**, and **C** and checking when the values are nonzero will show that the implicit function theorem gives us the same answers as before.

Finally, it is important to realize that the implicit function theorem says nothing about the **existence** of solutions to equations

$$f(\mathbf{x},\mathbf{y}) = \mathbf{0},$$

but rather is typically used to assert that the solution set, **if** it exists, can be parameterized locally by

$$\mathbf{x} \to (\mathbf{x}, g(\mathbf{x}))$$

using some function g.

**4.4.8. Example.** Consider the set of points **C** in $\mathbf{R}^3$ defined by the equations

$$\begin{aligned} x^2 + yz + 1 &= 0 \\ x^2 - y^2 + z^2 + 2 &= 0. \end{aligned}$$

To show that a neighborhood of the point $\mathbf{p} = (1,2,-1)$ on **C** is a curve that can be parameterized by a function $\gamma(x)$.

**Solution.** Define a function

$$f: \mathbf{R} \times \mathbf{R}^2 \to \mathbf{R}^2$$

by

$$f(x,y,z) = \left(x^2 + yz + 1, x^2 - y^2 + z^2 + 2\right).$$

Then the matrix of partials

$$M = \begin{pmatrix} \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \end{pmatrix} = \begin{pmatrix} z & y \\ -y & 2z \end{pmatrix}$$

is nonsingular at $\mathbf{p}$. The set $\mathbf{C}$ can now be parameterized by the function

$$\gamma(x) = (x, g(x)),$$

where $g: \mathbf{A} \to \mathbf{R}^2$ is the function defined on a neighborhood $\mathbf{A}$ of 1 in $\mathbf{R}$ guaranteed to exist by the implicit function theorem.

## 4.5 Critical Points

In this section we review some basic results about maxima and minima of functions, in particular, for functions of one or two variables.

**4.5.1. Theorem**. Let $\mathbf{X} \subseteq \mathbf{R}^n$ and let $f: \mathbf{X} \to \mathbf{R}$ be a continuous function. If $\mathbf{X}$ is compact, then f assumes both a minimum and maximum value on $\mathbf{X}$, that is, there are $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbf{X}$ so that

$$f(\mathbf{p}_1) \le f(\mathbf{p}) \le f(\mathbf{p}_2)$$

for all $\mathbf{p}$ in $\mathbf{X}$.

**Proof.** By Theorem 4.2.11, the set $\mathbf{Y} = f(\mathbf{X})$ is compact. By Theorem 4.2.4, $\mathbf{Y}$ is closed and bounded, so that both inf $\mathbf{Y}$ and sup $\mathbf{Y}$ belong to $\mathbf{Y}$. Choose any $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbf{X}$ with inf $\mathbf{Y} = f(\mathbf{p}_1)$ and sup $\mathbf{Y} = f(\mathbf{p}_2)$.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$ be an open set and let $f: \mathbf{X} \to \mathbf{R}$ be a differentiable function. Let $\mathbf{p} \in \mathbf{X}$. If $Df(\mathbf{p}) = 0$, then $\mathbf{p}$ is called a *critical point* of f and $f(\mathbf{p})$ is called a *critical value*.

Note that from a practical point of view, to check whether a point $\mathbf{p}$ is a critical point of f one simply checks if all the partials of f vanish at $\mathbf{p}$.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$ and let $f: \mathbf{X} \to \mathbf{R}$. The function f is said to have a *local maximum* at a point $\mathbf{p}$ in $\mathbf{X}$ if $f(\mathbf{q}) \le f(\mathbf{p})$ for all $\mathbf{q}$ in a neighborhood of $\mathbf{p}$. The function f is said to have a *local minimum* at $\mathbf{p}$ if $f(\mathbf{q}) \ge f(\mathbf{p})$ for all $\mathbf{q}$ in a neighborhood of $\mathbf{p}$. A point $\mathbf{p}$ is a *local extremum* if it is either a local maximum or local minimum. A point $\mathbf{p}$ in $\mathbf{X}$ is a (global) *maximum* for f if $f(\mathbf{q}) \le f(\mathbf{p})$ for all $\mathbf{q}$ in $\mathbf{X}$. The point $\mathbf{p}$ is a (global) *minimum* for f if $f(\mathbf{q}) \ge f(\mathbf{p})$ for all $\mathbf{q}$ in $\mathbf{X}$. An *extremum* for f is either a maximum or minimum for f.

**4.5.2. Theorem.** Let $\mathbf{X} \subseteq \mathbf{R}^n$ and let $f: \mathbf{X} \to \mathbf{R}$ be a differentiable function. If a point $\mathbf{p}$ in the interior of $\mathbf{X}$ is a relative extremum for f, then $\mathbf{p}$ is a critical point of f.

**Proof.** We give a proof for the case where $\mathbf{p}$ is a local maximum. In that case, $f(\mathbf{q}) \leq \mathbf{p}$ for all points $\mathbf{q}$ sufficiently close to $\mathbf{p}$. The definition of the directional derivative implies that $D_{\mathbf{v}}f(\mathbf{p}) \leq 0$ for all directions $\mathbf{v}$. In particular, $D_{\mathbf{v}}f(\mathbf{p}) \leq 0$ and $D_{-\mathbf{v}}f(\mathbf{p}) \leq 0$. This and Proposition 4.3.18(1) clearly imply that $D_{\mathbf{v}}f(\mathbf{p}) = 0$, so that $\nabla f(\mathbf{p}) = \mathbf{0}$.

**4.5.3. Theorem.** Let $f:[a,b] \to \mathbf{R}$ be a $C^2$ function and assume that c in (a,b) is a critical point of f.

(1) If $f''(c) < 0$, then c is a local maximum for f.
(2) If $f''(c) > 0$, then c is a local minimum for f.
(3) If $f''(c) = 0$, then nothing can be concluded from this test.

**Proof.** The easiest way to prove (1) and (2) is to use the Taylor expansion for f. See [Buck78]. The canonical examples for (1) and (2) are the functions $-x^2$ and $x^2$, respectively. Their graphs are shown in Figure 4.12(a). To prove (3), simply consider the functions $x^4$, $-x^4$, and $x^3$ (Figure 4.12(b)). This finishes the proof of Theorem 4.5.3.

The graph of the function $x^3$ shows another property of graphs.

**Definition.** A point c is called an *inflection point* of a function f(x) if there is an $\varepsilon > 0$, so that either

$$f''(x) < 0 \text{ for } x \in (c-\varepsilon, c) \quad \text{and} \quad f''(x) > 0 \text{ for } x \in (c, c+\varepsilon)$$

or

$$f''(x) > 0 \text{ for } x \in (c-\varepsilon, c) \quad \text{and} \quad f''(x) < 0 \text{ for } x \in (c, c+\varepsilon).$$

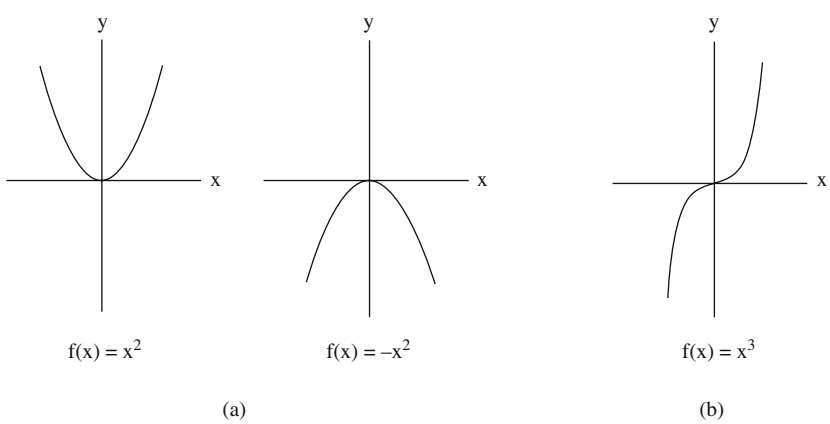



**Figure 4.12.** Some canonical cases of Theorem 4.5.3.

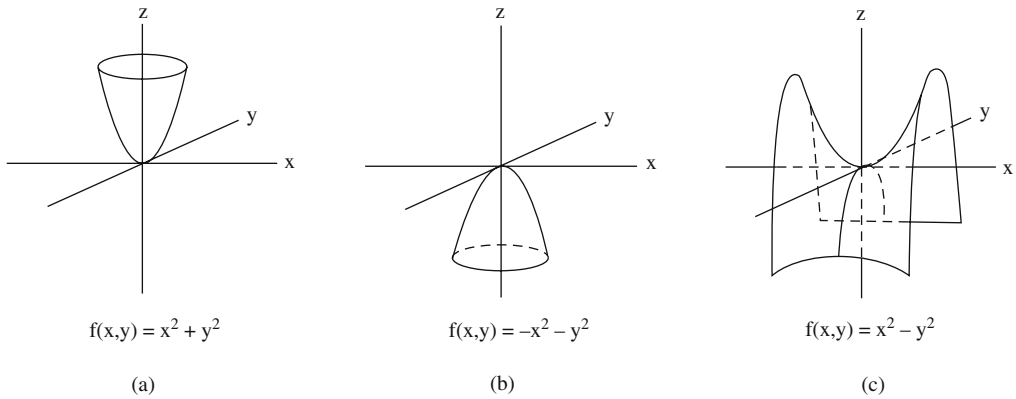



**Figure 4.13.** Extrema for functions of two variables.

An inflection point is a place where the "concavity" of the graph of a function changes from "concave upward" to "concave downward" or vice versa. The derivative does **not** have to vanish at an inflection point. The origin is an inflection point of the function $x^3$. See Figure 4.12(b).

Theorem 4.5.3 is not the end of the story for finding extrema for functions of one variable. One must still check a function on its endpoints. For example, the function $f(x) = x$ defined on $[0,1]$ does not have any critical points on $[0,1]$ but obviously has a maximum and a minimum, but these come on the endpoints of the interval.

Next, we want to extend the results about functions of one variable to functions of two variables. The canonical examples are the functions $-x^2 - y^2$ and $x^2 + y^2$, which have a local maximum and a local minimum, respectively, at the origin (see Figure 4.13(a) and (b)), but there is one more possibility. Consider the function

$$f(x,y) = x^2 - y^2. \tag{4.13}$$

See Figure 4.13(c). Although the origin is a critical point for f, it is a minimum along the x-axis and a maximum along the y-axis, that is, it is not a relative extremum but a "saddle point."

**Definition.** A critical point of a function that is not a relative extremum is called a *saddle point*.

**4.5.4. Theorem.** Let $\mathbf{X} \subseteq \mathbf{R}^2$ and let $f: \mathbf{X} \to \mathbf{R}$ be a $C^2$ function. Assume that $\mathbf{p}$ is a critical point for f that lies in the interior of $\mathbf{X}$. Let

$$D = \left(f_{xx}f_{yy} - f_{xy}^2\right)\big|_{\mathbf{p}}.$$

(1) If $D > 0$, then $\mathbf{p}$ is a relative extremum for f that is a relative maximum if $f_{xx}(\mathbf{p}) < 0$ and a relative minimum if $f_{xx}(\mathbf{p}) > 0$.
(2) If $D < 0$, then $\mathbf{p}$ is a saddle point.
(3) If $D = 0$, then nothing can be concluded from this test.

**Proof.** The Taylor expansion for functions of two variables can be used to prove (1) and (2). See [Buck78]. It is easy to check that the functions in Figure 4.13(a) and (b) fall into case (1) and that the function in Figure 4.13(c) (equation (4.13)) falls into case (2). To prove (3), consider the functions

(a) $f(x,y) = x^4 + y^4$
(b) $g(x,y) = -(x^4 + y^4)$
(c) $h(x,y) = x^2 - 5xy^2 + 4y^4 = (x - y^2)(x - 4y^2)$

Each of these functions has $D = 0$. The function f has a minimum at (0,0) and g has a maximum at (0,0). The function h has a saddle point at (0,0). See Figure 4.14(a). The "+" and "–" in the figure indicate the regions where the function is strictly positive or strictly negative, respectively. On the other hand, note that h(x,y) has a local minimum along any line through the origin. To see this, define

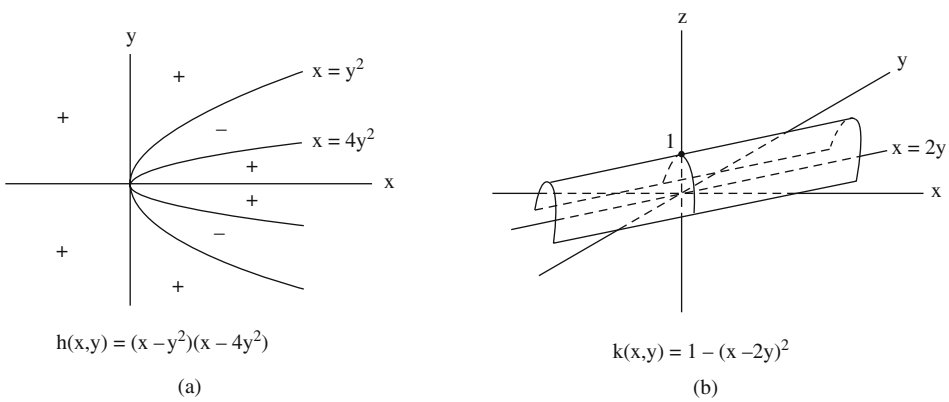
$$s(x) = h(x,mx) = \left(x - m^2x^2\right)\left(x - 4m^2x^2\right) = x^2 - 5m^2x^3 + 4m^4x^4.$$

The function s(x) describes h(x,y) along the line $y = mx$. A direct computation shows that $s'(0) = 0$ and $s''(0) = 2$, proving the claim. Along the y-axis we have that $h(0,y) = 4y^4$, and so we also have a local minimum at 0. This finishes the proof of Theorem 4.5.4, but before we move on, consider another interesting function that has $D = 0$, namely,

$$k(x,y) = 1 - x^2 + 4xy - 4y^2 = 1 - (x - 2y)^2.$$

See Figure 4.14(b). Here we have a function that is constant along every line of the form $x - 2y = c$. Every point on the line $x - 2y = 0$ is a critical point.

As in the case of functions of one variable, there is more to finding relative extrema of functions of two variables than just Theorem 4.5.4. We must always check the function on the boundary of its domain separately. This basically reduces to finding the relative extrema of a function of one variable, a problem we have already handled.




$h(x,y) = (x - y^2)(x - 4y^2)$ (a)

$k(x,y) = 1 - (x - 2y)^2$ (b)

**Figure 4.14.** Case (3) of Theorem 4.5.4.

**4.5.5. Example.** To analyze the extrema of the function

$$f(x,y) = x^2y^2$$

on the unit disk.

**Solution.** Looking for critical points alone will find its minimum at (0,0) but not its maximum, which lies on the boundary of the unit disk. To find that, make the substitutions $x = \cos\theta$ and $y = \sin\theta$ and define a function

$$g(\theta) = \cos^2\theta\sin^2\theta = (1/4)\sin^2 2\theta.$$

We only need to find the maxima of the function $g(\theta)$ for $\theta \in [0,2\pi]$. Now

$$g'(\theta) = \sin 2\theta\cos 2\theta = (1/2)\sin 4\theta.$$

Therefore, $g'(\theta) = 0$ implies that $\theta = k(\pi/4)$, for $k = 0,1,\ldots,7$. The second derivative test shows that g takes on its maximum values when $\theta = k(\pi/4)$, for $k = 1,3,5$, and 7.

Another common type of extremum problem is finding extrema subject to certain constraints.

**4.5.6. Example.** Maximize the function $f(x,y) = xy$ subject to the condition that $x^2 + y^2 = 1$.

The straighforward way to solve this problem would be to solve for y in the constraint equation and then substitute this into the formula for f thereby reducing the problem to a problem about extrema of functions of one variable. Unfortunately, this is not always feasible since the constraints may be much more complicated. We therefore want to briefly mention another popular approach to these types of problems. We begin with two facts that motivate the approach. We shall see the second, Proposition 4.5.8, again in Chapter 8 when we discuss tangent vectors to surfaces.

**4.5.7. Proposition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$ be an open set and assume that $f:\mathbf{X} \to \mathbf{R}$ and $\gamma:[a,b] \to \mathbf{X}$ are differentiable functions. If $\mathbf{p} = \gamma(c)$ is an extremum of f along $\gamma$ for some c in (a,b), then $\nabla f(\mathbf{p}) \bullet \gamma'(c) = 0$.

**Proof.** Consider the function $g(t) = f(\gamma(t))$. Since c is an extremum for g, it follows that $g'(c) = 0$. The result now follows from the chain rule, which says that

$$g'(t) = \nabla f(\gamma(t)) \bullet \gamma'(t).$$

**4.5.8. Proposition.** Let $g:\mathbf{R}^n \to \mathbf{R}$ be a differentiable function. If $\gamma:[a,b] \to g^{-1}(0)$ is a differentiable function, then $\nabla g \bullet \gamma' = 0$.

**Proof.** Consider the function $h(t) = g(\gamma(t))$. By hypothesis, $h(t) = 0$ for all t. Therefore, $h'(t) = 0$ and the chain rule gives the result.

Let us now restate our new extremum problem as follows:

**The constrained extremum problem:** Find the extrema of a function $f(x_1, \ldots, x_n)$ subject to a constraint $g(x_1, \ldots, x_n) = 0$.

Note first that the set $\mathbf{Z} = g^{-1}(0)$ of zeros of g is intuitively an $(n - 1)$-dimensional space with unique normal line. Therefore, Propositions 4.5.7 and 4.5.8 suggest that $\nabla f$ and $\nabla g$ must be parallel on $\mathbf{Z}$ since they are vectors that are both perpendicular to all curves in $\mathbf{Z}$. This observation leads to the following:

**The method of Lagrange multipliers:** Find the extrema of f subject to a constraint g by solving the equation

$$\nabla(f - \lambda g) = 0.$$

for $\lambda$.

We shall demonstrate the use of the Lagrange method with two examples.

**Solution to Example 4.5.6.** If

$$F(x, y) = xy - \lambda\left(x^2 + y^2 - 1\right),$$

then

$$\frac{\partial F}{\partial x} = y - 2\lambda x \quad \text{and} \quad \frac{\partial F}{\partial y} = x - 2\lambda y.$$

Setting $\partial F/\partial x$ to zero, implies $\lambda = y/(2x)$. If we substitute this $\lambda$ into the equation $\partial F/\partial y = 0$ and solve for x, then we will get that $x = \pm y$. Now substitute $\pm y$ for x in the constraint equation and solve for y. It follows that local extrema occur at

$$(x, y) = \left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right).$$

**4.5.9. Example.** To show that of all the triangles inscribed in a fixed circle, the equilateral triangle has the largest perimeter.

**Solution.** See Figure 4.15. The length of the side subtended by angle $\alpha$ is $2R \sin(\alpha/2)$. A similar formula holds for the other sides. Therefore,

$$f(\alpha, \beta, \gamma) = 2R\left(\sin\frac{\alpha}{2} + \sin\frac{\beta}{2} + \sin\frac{\gamma}{2}\right)$$

is the formula for the perimeter with the constraint

$$g(\alpha, \beta, \gamma) = \alpha + \beta + \gamma - 2\pi = 0.$$

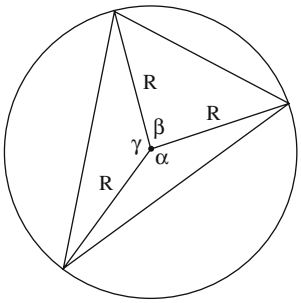



**Figure 4.15.** Inscribing a triangle in a circle.

If

$$F(\alpha,\beta,\gamma) = f(\alpha,\beta,\gamma) - \lambda g(\alpha,\beta,\gamma),$$

then

$$\frac{\partial F}{\partial \alpha} = R\cos\frac{\alpha}{2} - \lambda$$
$$\frac{\partial F}{\partial \beta} = R\cos\frac{\beta}{2} - \lambda$$
$$\frac{\partial F}{\partial \gamma} = R\cos\frac{\gamma}{2} - \lambda.$$

Setting the partials to zero, it follows that

$$\cos\frac{\alpha}{2} = \cos\frac{\beta}{2} = \cos\frac{\gamma}{2}.$$

In other words, $\alpha = \beta = \gamma$.

Next, we describe an application of directional derivatives to the extremum problem and an application of Proposition 4.5.8.

**4.5.10. Theorem.** The greatest rate of change of a function $f: \mathbf{R}^n \to \mathbf{R}$ at a point $\mathbf{p}$ takes place in the direction of and has the magnitude of the vector $\nabla f(\mathbf{p})$.

**Proof.** We are looking for the direction in which the directional derivative has a maximum. Let $\mathbf{u}$ be a unit vector. From Proposition 4.3.18(1) we know that

$$D_{\mathbf{u}} f = \nabla f \bullet \mathbf{u} = |\nabla f||\mathbf{u}|\cos\theta = |\nabla f|\cos\theta,$$

where $\theta$ is the angle between $\mathbf{u}$ and $\nabla f$. Clearly, this value will be a maximum when $\theta$ is 0 or $\pi$.

**4.5.11. Theorem.** Given a differentiable function $f: \mathbf{R}^n \to \mathbf{R}$, to move along a contour $f(\mathbf{p}) = c$, one should move in a direction which is orthogonal to $\nabla f$.

**Proof.** Apply Proposition 4.5.8 to the function $g(\mathbf{p}) = f(\mathbf{p}) - c$. Note that $\nabla g = \nabla f$.

We finish this section with two applications about the existence and uniqueness of closest points between planes.

**4.5.12. Theorem.** Let $\mathbf{X}$ be a k-dimensional plane in $\mathbf{R}^n$. For any point $\mathbf{p}$ in $\mathbf{R}^n$ there is a **unique** point $\mathbf{x}$ in $\mathbf{X}$ that is closest to $\mathbf{p}$. The point $\mathbf{x}$ is defined by the condition that the vector $\mathbf{px}$ is orthogonal to $\mathbf{X}$.

**Proof.** Let $\mathbf{x}_0 \in \mathbf{X}$ and assume that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are an orthonormal basis for $\mathbf{X}$. Parameterize the points of $\mathbf{X}$ with the function $\varphi : \mathbf{R}^k \to \mathbf{X}$ defined by

$$\varphi(t_1, t_2, \ldots, t_k) = \mathbf{x}_0 + t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \ldots + t_k\mathbf{u}_k.$$

Define differentiable functions $\sigma : \mathbf{R}^n \to \mathbf{R}$ and $d : \mathbf{R}^k \to \mathbf{R}$ by

$$\sigma(\mathbf{x}) = \mathbf{xp} \bullet \mathbf{xp}$$

and

$$d(\mathbf{t}) = \sigma(\varphi(\mathbf{t})) = (\mathbf{p} - \varphi(\mathbf{t})) \bullet (\mathbf{p} - \varphi(\mathbf{t})).$$

Clearly, finding a point $\mathbf{x}$ in $\mathbf{X}$ that is closest to $\mathbf{p}$ is equivalent to finding a minimum of the function d.

Our first observation is that d does achieve a minimum and that this minimum is a critical point of d. This follows from Theorems 4.5.1 and 4.5.2. The domain of d, $\mathbf{R}^k$, is of course not a compact set, but we can apply the theorems to the function d restricted to some large closed disk in $\mathbf{R}^k$ with the property that d is larger at every point on the boundary of that disk than at some point on its interior. This will guarantee that a relative extremum will occur in the interior of the disk. Such a disk clearly exists because $d(\mathbf{t})$ goes to infinity as $|\mathbf{t}|$ goes to infinity.

Next, we show that d has a unique minimum, one defined by the stated orthogonality condition. The chain rule applied to d gives that

$$\frac{\partial d}{\partial t_i}(\mathbf{t}) = -2\mathbf{u}_i \bullet (\mathbf{p} - \varphi(\mathbf{t})).$$

Since the critical points $\mathbf{t}$ of d are the points where all of these partials vanish, we see that solving for those points is equivalent to solving for those points $\mathbf{x} = \varphi(\mathbf{t})$ in $\mathbf{X}$ satisfying $\mathbf{u}_i \bullet \mathbf{xp} = 0$, for all i, which shows that $\mathbf{xp}$ must be orthogonal to $\mathbf{X}$.

Finally, we need to show that a point of $\mathbf{X}$ defined by such orthogonality conditions is unique. See Figure 4.16. Let $\mathbf{x}' = \varphi(\mathbf{t}')$ and assume that $\mathbf{u}_i \bullet \mathbf{x}'\mathbf{p} = 0$, for all i. This would imply that $\mathbf{u}_i \bullet \mathbf{xx}' = 0$, for all i. Since $\mathbf{xx}'$ is a vector in the plane $\mathbf{X}$ and the $\mathbf{u}_i$ form a basis for $\mathbf{X}$, we must have that $\mathbf{xx}' = \mathbf{0}$, that is, $\mathbf{x} = \mathbf{x}'$. The theorem is proved.

Theorem 4.5.12 is a special case of the next theorem. We sketched the proof in this special case because, being simpler, it brought out more clearly the essential steps in the proof of these types of theorems.

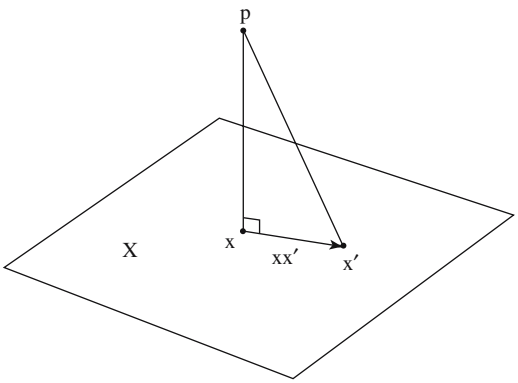



**Figure 4.16.** Uniqueness of closest point **x** in plane.

**4.5.13. Theorem.** If **X** and **Y** are transverse planes in $\mathbf{R}^n$ with

$$\dim \mathbf{X} + \dim \mathbf{Y} \le n,$$

then there are **unique** points $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$, so that

$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) = |\mathbf{xy}| = \operatorname{dist}(\mathbf{X}, \mathbf{Y}).$$

The points **x** and **y** are defined by the condition that the vector **xy** is orthogonal to both **X** and **Y**.

**Proof.** It is easy to show that proving the theorem reduces to proving the following two cases:

**Case 1:** $\dim \mathbf{X} + \dim \mathbf{Y} = n$

**Case 2:** $\dim \mathbf{X} + \dim \mathbf{Y} = n - 1$

In Case 1 the planes intersect in a single point. We sketch the proof for Case 2 and leave the rest to the reader (Exercise 4.5.4). Case 2 applies to skew lines in $\mathbf{R}^3$, for example. The fact that there are points **x** and **y** at which the distance between **X** and **Y** is minimized and the fact that **xy** is orthogonal to **X** and **Y** is proved just like in Theorem 4.5.12 by parameterizing points of **X** and **Y** via tuples **s** and **t**, expressing the distance between points of **X** and **Y** as a function d(**s**,**t**) of the variables **s** and **t**, and looking for the critical points of that function by setting the partial derivatives of d(**s**,**t**) to zero. It remains to show that the orthogonality condition defines **x** and **y** uniquely.

Assume that there are other points $\mathbf{x}' \in \mathbf{X}$ and $\mathbf{y}' \in \mathbf{Y}$ with the property that the vector $\mathbf{x}'\mathbf{y}'$ is orthogonal to both **X** and **Y**. See Figure 4.17. Our hypothesis about the dimensions of **X** and **Y** implies that all vectors that are orthogonal to **both X** and **Y** are multiples of each other. Therefore, **xy** and $\mathbf{x}'\mathbf{y}'$ are parallel. If $(\mathbf{x},\mathbf{y}) \ne (\mathbf{x}',\mathbf{y}')$, then the points **x**, **y**, $\mathbf{x}'$, and $\mathbf{y}'$ lie in a two-dimensional plane **Z**. Assume that $\mathbf{x} \ne \mathbf{x}'$ and $\mathbf{y} \ne \mathbf{y}'$ and consider the line **L** through **x** and $\mathbf{x}'$ and the line $\mathbf{L}'$ through **y** and $\mathbf{y}'$. (The

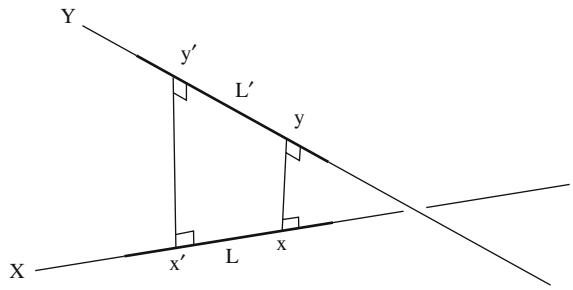



**Figure 4.17.** Uniqueness of closest points **x** and **y**.

special cases where either $\mathbf{x} = \mathbf{x}'$ or $\mathbf{y} = \mathbf{y}'$ is left as an exercise for the reader.) Now the lines **L** and **L′** are parallel in **Z** because they are both orthogonal to the vector **xy**. On the other hand, **L** and **L′** lie in **X** and **Y**, respectively, which would imply that **X** and **Y** would have a common basis vector. This contradicts the fact that **X** and **Y** are transverse planes. This contradiction proves the theorem for Case 2.

## 4.6 Morse Theory

The object of this section is to analyze functions in a neighborhood of a critical point. The results will have important applications in Chapter 8.

Let $\mathbf{X} \subseteq \mathbf{R}^n$ be an open set and let $f: \mathbf{X} \to \mathbf{R}$ be a $C^\infty$ function. Let $\mathbf{p} \in \mathbf{X}$.

**Definition.** The $n \times n$ matrix of second partials

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{p})\right)$$

is called the *Hessian matrix* of f at **p**. Its determinant is called the *Hessian* of f at **p**.

**Definition.** A critical point **p** of f is called a *nondegenerate critical point* if the Hessian matrix of f at **p** is nonsingular. Otherwise, it is called *degenerate*.

**4.6.1. Examples.**

(1) The origin is a nondegenerate critical point for $f(x) = x^2$, $g(x,y) = x^2 + y^2$, and $h(x,y) = x^2 - y^2$.
(2) The origin is a degenerate critical point for $f(x) = x^3$ and $g(x,y) = x^2 + y^3$.

**4.6.2. Lemma.** Let f be a real-valued $C^\infty$ function defined in a convex neighborhood **V** of **0** in $\mathbf{R}^n$ with $f(\mathbf{0}) = 0$. Then

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i g_i(x_1, x_2, \ldots, x_n),$$

where the $g_i$ are $C^\infty$ functions defined on $\mathbf{V}$ with $g_i(\mathbf{0}) = \partial f/\partial x_i(\mathbf{0})$.

**Proof.** Consider the function $t \to f(tx_1, tx_2, \ldots, tx_n)$. By the Fundamental Theorem of Calculus

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \int_0^1 \frac{df}{dt}(tx_1, tx_2, \ldots, tx_n)dt \\ &= \int_0^1 \left[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(tx_1, tx_2, \ldots, tx_n)x_i\right]dt. \end{aligned}$$

We can therefore define the functions $g_i$ by

$$g_i(x_1, x_2, \ldots, x_n) = \int_0^1 \frac{\partial f}{\partial x_i}(tx_1, tx_2, \ldots, tx_n)dt$$

and the lemma is proved. (One needs $C^\infty$ here because if f is only $C^r$, then the $g_i$ will not be $C^r$ in general.)

The main result of this section is the next theorem.

**4.6.3. Theorem.** (The Morse Lemma) Let f be a real-valued $C^\infty$ function defined in a neighborhood of a point $\mathbf{p}$ in $\mathbf{R}^n$. If $\mathbf{p}$ is a nondegenerate critical point for f, then there is a local diffeomorphism $\varphi$ from a neighborhood $\mathbf{V}$ of $\mathbf{p}$ onto a neighborhood $\mathbf{U}$ of $\mathbf{0}$ in $\mathbf{R}^n$ so that

$$(f \circ \varphi^{-1})(y_1, y_2, \ldots, y_n) = f(\mathbf{p}) - y_1^2 - \ldots - y_k^2 + y_{k+1}^2 + \ldots + y_n^2$$

on $\mathbf{U}$. The integer k is well defined and called the *index* of $\mathbf{p}$.

**Proof.** Without loss of generality assume that $\mathbf{p} = \mathbf{0}$ and $f(\mathbf{p}) = \mathbf{0}$. By Lemma 4.6.2 we can write f as

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^n x_i g_i(x_1, x_2, \ldots, x_n).$$

Since $\mathbf{p}$ is a critical point of f,

$$g_i(\mathbf{0}) = \frac{\partial f}{\partial x_i}(\mathbf{0}) = 0.$$

This means that we can apply Lemma 4.6.2 to the $g_i$ to get

$$g_i(x_1, x_2, \ldots, x_n) = \sum_{j=1}^n x_j h_{ij}(x_1, x_2, \ldots, x_n).$$

for some $C^\infty$ functions $h_{ij}$. Hence

$$f(x_1, x_2, \ldots, x_n) = \sum_{i,j=1}^{n} x_i x_j h_{ij}(x_1, x_2, \ldots, x_n).$$

By replacing $h_{ij}$ by $(1/2)(h_{ij} + h_{ji})$, if necessary, we may assume that $h_{ij} = h_{ji}$ in some neighborhood of $\mathbf{0}$. Furthermore, the matrix $(h_{ij}(\mathbf{0}))$ is just the Hessian matrix of f. Since this matrix is assumed to be nonsingular, we can copy the diagonalization proof for quadratic forms given in Theorem 1.9.11 (see also [Miln63]) to finish the result.

**4.6.4. Corollary.** Nondegenerate critical points of functions are isolated.

The Morse Lemma and its corollary show that one has a good understanding of what the graph of functions look like near a nondegenerate critical point. The situation is much more complicated in the degenerate case.

**4.6.5. Example.** Consider the functions

(1) $f(x) = e^{-1/x^2}\sin^2(1/x)$ (Figure 4.18(a))
(2) $f(x,y) = x^2$ (Figure 4.18(b))
(3) $f(x,y) = x^2y^2$ (Figure 4.18(c))

All these functions have the origin as a nonisolated degenerate critical point. The function in (1) has a sequence of nondegenerate critical points converging to the origin. All the points on the y-axis are degenerate critical points for the function in (2). All the points on both the x- and y-axis are degenerate critical points of the function in (3).

We shall see later in Chapter 8 that one can tell a lot about the topology of a space from the nondegenerate critical points of functions defined on them.

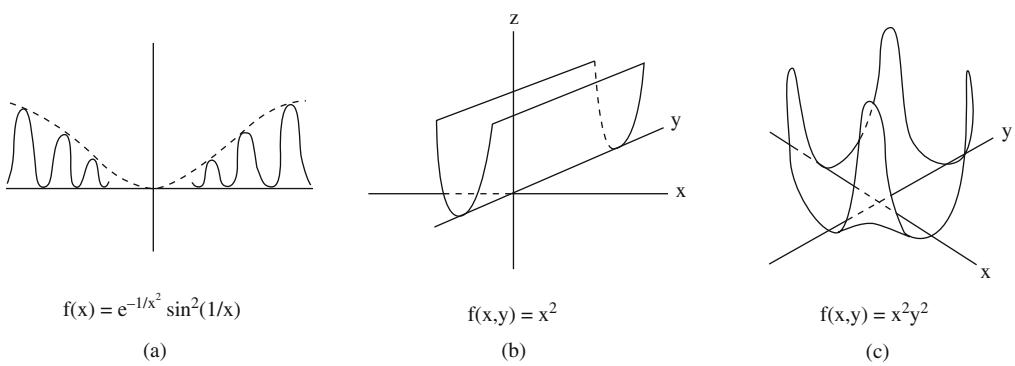



**Figure 4.18.** Examples of nonisolated degenerate critical points.

## 4.7 Zeros of Functions

Many problems can be reduced to the problem of finding the zeros or roots of a function. This section discusses some basic approaches to solving this problem.

**The Problem:** Given a function $f: \mathbf{R}^n \to \mathbf{R}^m$, find a solution to

$$f(\mathbf{x}) = \mathbf{0}. \tag{4.14}$$

**The Newton-Raphson method:** We begin with the case where n = m = 1. Pick a guess $x_0$ for a root. If this is not correct, then from the Taylor expansion formula we know that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \text{higher-order terms.}$$

Forgetting the higher-order terms means that, as an approximation, we are looking for an x so that

$$0 = f(x_0) + f'(x_0)(x - x_0).$$

In other words,

$$x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

We use this x as the next guess at a solution to f(x) = 0. If we still do not have a root we repeat this process, thereby generating a sequence of points $x_0, x_1, x_2, \ldots,$ where in general,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

This sequence hopefully converges to a root. Figure 4.19 shows what is going on geometrically. If $x_k$ is not a root, then find the intersection of the tangent line to the graph of f(x) at $(x_k, f(x_k))$ with the x-axis. This point becomes our next guess $x_{k+1}$.

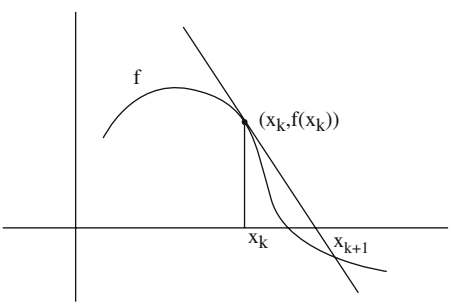



**Figure 4.19.** The Newton-Raphson method.

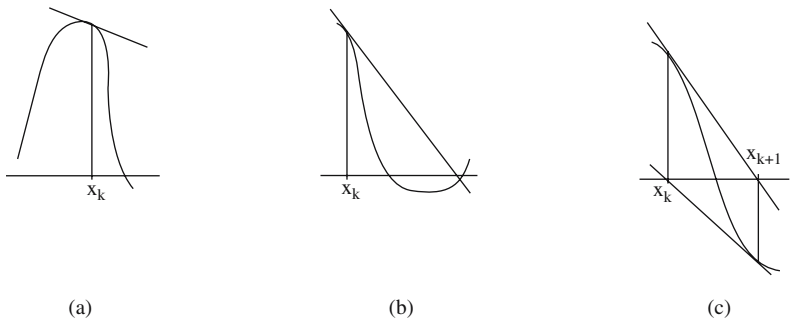



**Figure 4.20.** Problems with the Newton-Raphson method.

The Newton-Raphson method is pretty good. There are theorems about when and how fast the sequence of points it generates converges (it basically has quadratic convergence) and upper bounds for errors. (See any text on numerical analysis.) There are some well-known problems however:

(1) Near critical points, even if we ignore the problem of numerical instability since we are dividing by a very small quantity, points of the sequence may move far away from the actual root (see Figure 4.20(a)).
(2) The sequence of points can home in on a more distant and "wrong" root (see Figure 4.20(b)).
(3) The points of the sequence may oscillate and not converge (see Figure 4.20(c)).
(4) Near multiple roots (where $f'(x)$ approaches 0 as $f(x)$ does), we may have slow convergence.

More sophisticated methods that overcome some of these problems are known, nevertheless, the method is widely used because it is so simple to implement. It seems to converge very rapidly in practice and one often has very high accuracy for the root after only several iterations. In any case though, it always helps if one knows something about the function f. If one can make a good initial estimate, then one is usually in good shape. Of course, choosing a good initial estimate is often a major problem with using the method. Some bounds for the location of zeros are known. The reader will find more information in most books on numerical methods. For an interesting history of how the method got its name see [Alex96]. Finally, note that what we have just said about real functions also applies to complex functions $f: \mathbf{C} \to \mathbf{C}$.

The Newton-Raphson method generalizes to higher dimensions and functions of several variables because there is a similar Taylor expansion

$$f(\mathbf{x}) = f(\mathbf{x}_k) + Df(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) + \text{higher-order terms.}$$

In other words, the only difference between what we have now and what we had before is that now we need to solve a **system** of linear equations

$$f(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)f'(\mathbf{x}_k) = \mathbf{0} \tag{4.15}$$

for **x**. We shall work out the case where n is 2 as an example.

**4.7.1. Example.** To find a root to the system of equations

$$f(x,y) = 0$$
$$g(x,y) = 0.$$

**Solution.** We define a sequence of points $(x_k,y_k)$ that converge to the root. Suppose that we already have defined the kth point $(x_k,y_k)$. The Taylor expansions for f and g around $(x_k,y_k)$ are

$$f(x,y) = f(x_k,y_k) + f_x(x_k,y_k)(x-x_k) + f_y(x_k,y_k)(y-y_k) + \dots$$
$$g(x,y) = g(x_k,y_k) + g_x(x_k,y_k)(x-x_k) + g_y(x_k,y_k)(y-y_k) + \dots$$

Truncating higher order terms means that we want to solve

$$f_x(x_k,y_k)(x-x_k) + f_y(x_k,y_k)(y-y_k) = -f(x_k,y_k)$$
$$g_x(x_k,y_k)(x-x_k) + g_y(x_k,y_k)(y-y_k) = -g(x_k,y_k).$$

Cramer's rule implies that

$$x_{k+1} = x_k + \frac{-fg_y + gf_y}{J}$$
$$y_{k+1} = y_k + \frac{-gf_x + fg_x}{J},$$

where $J = f_xg_y - g_xf_y$ and all the partials are evaluated at $(x_k,y_k)$. If $(x_0,y_0)$ is an approximation to a root (a,b), then one can show that convergence to (a,b), if any, is quadratic. Conditions that are sufficient to guarantee such convergence assuming that $(x_0,y_0)$ is "close enough" to (a,b) are of the type:

(1) The derivatives of f and g up to order 2 are continuous and bounded in a neighborhood **U** of (a,b).
(2) The Jacobian J does not vanish in **U**.

See [ConD72].

Now methods for finding zeros of functions have a much wider application than simply that specific problem. Many other problems can be rephrased in those terms. Specifically, problems that involve finding the inverse of a function are often rephrased in terms of finding solutions to an equation of the form (4.14). As an example of that type of problem, suppose that one is given a function $p: \mathbf{R}^2 \to \mathbf{R}^3$, a point $\mathbf{q} \in \mathbf{R}^3$, and one wants to find values x and y so that $p(x,y) = \mathbf{q}$. This problem can be expressed in terms of finding a zero of the function

$$f(x,y) = \mathbf{q} - p(x,y). \tag{4.16}$$

The Newton-Raphson approach would start with a guess $\mathbf{z}_0 = (x_0,y_0)$ and then, using equation (4.15), generate a sequence of points $\mathbf{z}_k = (x_k,y_k)$ satisfying

$$(\mathbf{z}_{k+1} - \mathbf{z}_k)p'(\mathbf{z}_k) = \mathbf{q} - p(\mathbf{z}_k), \tag{4.17}$$

where $p'(\mathbf{z})$ is the Jacobian matrix

$$p'(\mathbf{z}) = \begin{pmatrix} \dfrac{\partial p}{\partial x}(\mathbf{z}) \\ \dfrac{\partial p}{\partial y}(\mathbf{z}) \end{pmatrix}.$$

If $p'(\mathbf{z})$ had an inverse, the one could solve equation (4.17) for $\mathbf{z}_{k+1}$ and one would then have a nice formula for generating the sequence of points $\mathbf{z}_{k+1}$, which would hopefully converge to a solution to equation (4.16). Unfortunately, the $2 \times 3$ matrix $p'(\mathbf{z})$ obviously does not have an inverse, but one can use the Moore-Penrose inverse instead to get the equation

$$\mathbf{z}_{k+1} - \mathbf{z}_k = (\mathbf{p} - p(\mathbf{z}_k))\left(p'^T(\mathbf{z}_k)\left(p'(\mathbf{z}_k)p'^T(\mathbf{z}_k)\right)^{-1}\right). \tag{4.18}$$

For an application of this approach to determine the intersection of two surfaces see [AbdY96].

Next, how do we detect and deal with problems encountered by the Newton-Raphson method? In general, one case where we have problems is when the gradient of a function f is zero. In that case, we would really need to analyze the Hessian of f. To avoid some of these difficulties, many other iterative methods have been developed. The basic situation is the following: We are trying to generate a sequence of points $\mathbf{x}_k$, where

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k \mathbf{d}_k.$$

In other words, the approach involved

(1) choosing a direction of search $\mathbf{d}_k$, and
(2) carrying out a linear search in that direction (specified by $\lambda_k$).

This is a "hill-climbing"–type problem. Looking at it that way leads to a steepest descent method (Cauchy, 1847) which chooses

$$\mathbf{d}_k = \lambda_k \nabla f(\mathbf{x}_k), \quad \lambda_k > 0. \tag{4.19}$$

This method is recommended when $\mathbf{x}_0$ is far from the root or a local critical point and then to use the Newton-Raphson iteration on

$$\nabla f = \mathbf{0}$$

when one gets close.

Finally, in the above discussion we have assumed that f was defined everywhere. However, if f has a restricted domain **D**, then additional problems arise, namely, what do we do when the new point $\mathbf{x}_{k+1}$ falls outside this domain? One needs to "clip" the point somehow. A guideline for handling such a situation when the domain is a rectangle is the following:

> If our current guess is **on** the boundary of **D** and the next forces us to go outside, then one should suspect a critical point **on** the boundary and deal with this as a special case. One uses as initial guess the corners and center of the rectangle.

## 4.8 Integration

This section will sketch how one can define integrals for functions of several variables and state several of the most important theorems. Proofs are omitted and can be found in [Spiv65] or [Buck78].

Let $\mathbf{A} \subseteq \mathbf{R}^n$ and $f: \mathbf{A} \to \mathbf{R}$. To define the integral of f we shall follow the sequence of steps below:

(1) We define the integral for the case where **A** is a rectangular set by using upper and lower sums for partitions of that set.
(2) We prove that the integral exists if the points of discontinuity of f are a set of measure zero.
(3) We extend the definition of the integral to arbitrary bounded sets **A** whose boundaries are sets of measure zero.

**Definition.** A subset **A** of $\mathbf{R}^n$ of the form $[a_1,b_1] \times [a_2,b_2] \times \ldots \times [a_n,b_n]$ with $b_i > a_i$, $i = 1, 2, \ldots, n$, is called an *n-rectangle* or simply a *rectangle* if n is clear from the context. The boundary of an n-rectangle consists of 2n planar pieces called the *faces* of the n-rectangle. The *volume* of **A**, denoted by vol(**A**), is defined by

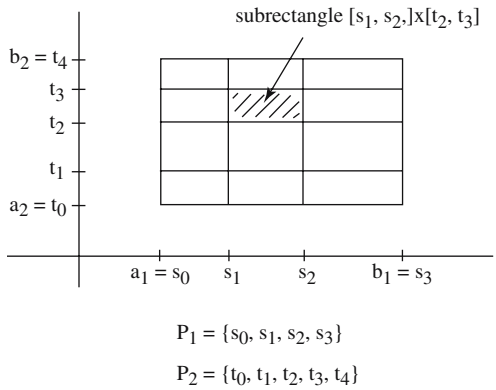
$$\mathrm{vol}(\mathbf{A}) = \prod_{i=1}^{n} (b_i - a_i).$$

If $b_1 - a_1 = b_2 - a_2 = \ldots = b_n - a_n$, then **A** is called an *n-dimensional cube* or *n-cube* or simply *cube*.

**Definition.** A *partition* of a rectangle $\mathbf{A} = [a_1,b_1] \times [a_2,b_2] \times \ldots \times [a_n,b_n]$ in $\mathbf{R}^n$ is a sequence $P = (P_1,P_2, \ldots ,P_n)$ where $P_i$ is a partition of the interval $[a_i,b_i]$. A *subrectangle* of the partition P is a rectangle $[c_1,d_1] \times [c_2,d_2] \times \ldots \times [c_n,d_n]$ where $[c_i,d_i]$ is a subinterval of the partition $P_i$. The *norm* of the partition P, denoted by $|P|$, is defined by

$$|P| = \max\{|P_i| \mid i = 1,2, \ldots ,n\}.$$

**Figure 4.21.** Subrectangles of a rectangle partition.




A *refinement* of the partition P is a partition $P' = (P_1', P_2', \ldots, P_n')$ of **A** where $P_i'$ is a refinement of $P_i$. See Fig 4.21.

**Definition.** Given a bounded function $f: \mathbf{X} \to \mathbf{R}$, define $\min_{\mathbf{X}}(f)$ and $\max_{\mathbf{X}}(f)$ by

$$\min_{\mathbf{X}}(f) = \inf\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}\},$$
$$\max_{\mathbf{X}}(f) = \sup\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}\}.$$

Let **A** be a rectangle in $\mathbf{R}^n$ and consider a function $f: \mathbf{A} \to \mathbf{R}$.

**Definition.** If P is a partition of **A**, then

$$L(f,P) = \sum_{\text{subrectangle } \mathbf{B} \text{ of } P} \min_{\mathbf{B}}(f)\mathrm{vol}(\mathbf{B})$$

and

$$U(f,P) = \sum_{\text{subrectangle } \mathbf{B} \text{ of } P} \max_{\mathbf{B}}(f)\mathrm{vol}(\mathbf{B})$$

are called the *lower* and *upper sum* for f over **A** with respect to P, respectively.

One can show that refining partitions increases the lower sums and decreases the upper sums and that upper sums are always larger than or equal to lower sums. This enables one to make the following definition.

**Definition.** The function f is said to be *integrable* on **A** if f is bounded on **A** and

$$\sup\{L(f,P) \mid P \text{ is a partition of } \mathbf{A}\} = \inf\{U(f,P) \mid P \text{ is a partition of } \mathbf{A}\}.$$

The common value is called the *integral* of f over **A** and is denoted by $\int_{\mathbf{A}} f$.

Having defined the integral, the next obvious task is to determine when it exists. One can show that continuous functions are integrable, but there is a more general condition. We shall not take the time to define the general concept of the measure ("volume") of a set. All we will need here is the following:

**Definition.** A subset $\mathbf{X}$ of $\mathbf{R}^n$ has *measure zero* if, for every $\varepsilon > 0$, it can be covered by a sequence of n-rectangles $\mathbf{A}_i$ so that

$$\sum_{i=1}^{\infty} \mathrm{vol}(\mathbf{A}_i) < \varepsilon.$$

It is easy to see that every finite set has measure zero. But countable sets $\{\mathbf{p}_i\}_{i=1}^{\infty}$ (like the rational numbers) also have measure zero because the ith one can be contained in a small rectangle of volume $\varepsilon/2^i$ and these volumes sum to $\varepsilon$. The boundary of "nice" sets, such as rectangles or closed disks, have measure zero. More generally, "k-dimensional" subsets of $\mathbf{R}^n$ have measure zero when $k < n$. For example, Figure 4.22 shows why a segment is a set of measure zero in $\mathbf{R}^2$. We can cover the segment $[(a,c),(b,d)]$ with n rectangles that have width $(b - a)/n$ and height $(d - c)/n$. The total area of these rectangles is

$$n\frac{b-a}{n}\frac{d-c}{n} = \frac{(b-a)(d-c)}{n}$$

and we can clearly make this area as small as desired by increasing n.

**4.8.1. Theorem.** Let $\mathbf{A}$ be rectangle in $\mathbf{R}^n$. A bounded function $f: \mathbf{A} \to \mathbf{R}$ is integrable if and only if the set

$$\{\ \mathbf{a} \in \mathbf{A} \mid f \text{ is discontinuous at } \mathbf{a}\ \}$$

has measure zero.

**Proof.** See [Spiv65] or [Buck78].

Theorem 4.8.1 is a good result for functions defined on rectangles, but what if the domain of a function is not a rectangle. Let $\mathbf{X}$ be a bounded subset of $\mathbf{R}^n$ and $f: \mathbf{X} \to \mathbf{R}$. Define $F: \mathbf{R}^n \to \mathbf{R}$ by

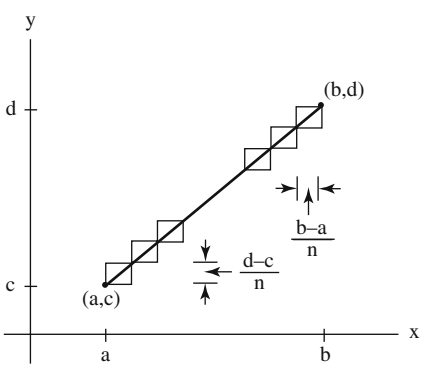



**Figure 4.22.** A segment is a set of measure zero in the plane.

$$\begin{aligned} F(\mathbf{x}) &= f(\mathbf{x}) \quad \text{if} \quad \mathbf{x} \in \mathbf{X} \\ &= 0 \quad \text{if} \quad \mathbf{x} \in \mathbf{R}^n - \mathbf{X}. \end{aligned}$$

and choose a rectangle **A** that contains **X**.

**Definition.** The *integral* of f over **X**, denoted by $\int_{\mathbf{X}} f$, is defined to be the integral $\int_{\mathbf{A}} F$, provided that integral is defined. If the integral of f exists, we say that f is *integrable* over **X**.

One can show that whether or not f is integrable on **X** is independent of the rectangle **A** that is chosen.

There is another way to phrase the problem of integrating over an arbitrary set.

**Definition.** Let $\mathbf{X} \subseteq \mathbf{R}^n$. The function

$$\chi_{\mathbf{X}} : \mathbf{R}^n \to \mathbf{R}$$

defined by

$$\begin{aligned} \chi_{\mathbf{X}}(\mathbf{x}) &= 1 \quad \text{if} \quad \mathbf{x} \in \mathbf{X} \\ &= 0 \quad \text{if} \quad \mathbf{x} \in \mathbf{R}^n - \mathbf{X}. \end{aligned} \tag{4.20}$$

is called the *characteristic function* of **X**.

If the function f above was actually defined over a rectangle **A**, but one simply wanted to integrate over a smaller set **X** in **A**, then an equivalent definition for the integral of f over **X** would be to define it to be $\int_{\mathbf{A}} f\chi_{\mathbf{X}}$ (if that integral exists).

**Definition.** A bounded set in $\mathbf{R}^n$ is said to be *Jordan-measurable* if its boundary is a set of measure 0.

**4.8.2. Theorem.** Let **X** be a Jordan-measurable subset of $\mathbf{R}^n$ and $f: \mathbf{X} \to \mathbf{R}$. If f is bounded on **X** and continuous on **X** except at possibly a set of measure zero, then f is integrable on **X**.

**Proof.** See [Spiv65] or [Buck78].

**4.8.3. Corollary.** If **X** is a bounded set in $\mathbf{R}^n$, then the characteristic function $\chi_{\mathbf{X}}$ of **X** is integrable over **X** if and only if **X** is a Jordan-measurable set.

We can use Theorem 4.8.2 (or its corollary) to define the volume of a set in $\mathbf{R}^n$.

**Definition.** Let **X** be a Jordan-measurable set in $\mathbf{R}^n$. Define the *volume* of **X**, denoted by vol(**X**), by

$$\text{vol}(\mathbf{X}) = \int_{\mathbf{X}} 1_{\mathbf{X}}.$$

Although we now know what it means for a function to be integrable, we still need a convenient way to compute the integral.

**4.8.4. Theorem.** (Fubini's Theorem) Let $\mathbf{A} \subset \mathbf{R}^s$ and $\mathbf{B} \subset \mathbf{R}^t$ be rectangles and assume that $f: \mathbf{A} \times \mathbf{B} \to \mathbf{R}$ is integrable. Define functions $g_{\mathbf{y}}: \mathbf{A} \to \mathbf{R}$ and $h_{\mathbf{x}}: \mathbf{B} \to \mathbf{R}$ by $g_{\mathbf{y}}(\mathbf{x}) = f(\mathbf{x},\mathbf{y})$ and $h_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x},\mathbf{y})$. If $g_{\mathbf{y}}(\mathbf{x})$ is integrable except possibly at a finite number of $\mathbf{y}$, then

$$\int_{\mathbf{A}\times\mathbf{B}} f = \int_{\mathbf{B}}\int_{\mathbf{A}} g_y = \int_{\mathbf{B}}\left(\int_{\mathbf{A}} f(\mathbf{x},\mathbf{y})\mathbf{dx}\right)\mathbf{dy}. \tag{4.21a}$$

Similarly, if $h_{\mathbf{x}}(\mathbf{y})$ is integrable except possibly at a finite number of $\mathbf{x}$, then

$$\int_{\mathbf{A}\times\mathbf{B}} f = \int_{\mathbf{A}}\int_{\mathbf{B}} h_x = \int_{\mathbf{A}}\left(\int_{\mathbf{B}} f(\mathbf{x},\mathbf{y})\mathbf{dy}\right)\mathbf{dx}. \tag{4.21b}$$

**Proof.** See [Spiv65] or [Buck78].

The integrals in equations (4.21a) and (4.21b) are called *iterated integrals*. One common situation where Fubini's theorem applies is when f is continuous.

It is worth restating Fubini's theorem in the special case where $f: [a,b] \times [c,d] \to \mathbf{R}$. Equations (4.21a) and (4.21b) then become

$$\begin{aligned}\int_{[a,b]\times[c,d]} f &= \int_a^b\left(\int_c^d f(x,y)dy\right)dx \\ &= \int_c^d\left(\int_a^b f(x,y)dx\right)dy.\end{aligned} \tag{4.22}$$

We can also apply Fubini's theorem to integrate over nonrectangular regions. We state one variant of those types of integrals. Suppose that $\mathbf{X} \subseteq [a,b] \times [c,d]$ and that $\mathbf{X}$ is bounded by the lines $x = a$, $y = b$, and the graphs of functions $\alpha(x)$ and $\beta(x)$ defined over $[a,b]$ with $\alpha(x) \le \beta(x)$. Then

$$\int_{\mathbf{X}} f = \int_a^b\left(\int_{\alpha(x)}^{\beta(x)} f(x,y)dy\right)dx. \tag{4.23}$$

Figure 4.23 shows the geometry. We are thinking of the right-hand side of equation (4.23) as an integral of the areas of the vertical slices.

**4.8.5. Example.** Let $\mathbf{X}$ be the region in the plane bounded by curves

$$x = 2, \quad x = 5, \quad y = 1 + (x-3)^2, \quad \text{and} \quad y = 6 - (x-4)^2.$$

See Figure 4.24. If $f(x,y) = y$, then

$$\int_{\mathbf{X}} f = \int_2^5\left(\int_{1+(x-3)^2}^{6-(x-4)^2} y\,dy\right)dx = \int_2^5 \frac{1}{2}\left[\left(6-(x-4)^2\right)^2 - \left(1+(x-3)^2\right)^2\right]dx = \frac{571}{10}.$$

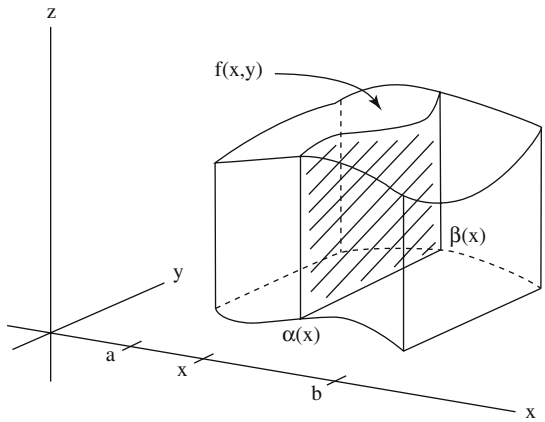



**Figure 4.23.** The iterated integral in equation (4.23).

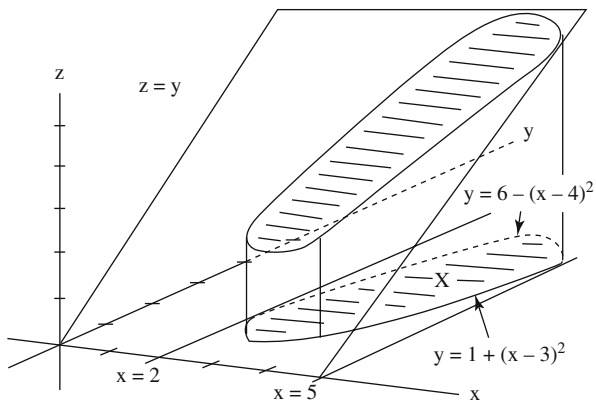



**Figure 4.24.** Example 4.8.5.

Finally, there is change of variable theorem for higher-dimensional integrals just like for real-valued functions of one variable.

**4.8.6. Theorem.** (The Generalized Change of Variables Theorem) Let $\mathbf{A} \subset \mathbf{R}^n$ be an open bounded subset and $g: \mathbf{A} \to \mathbf{R}^n$ a one-to-one and continuously differentiable function. Assume that the Jacobian matrix $g'$ has a nonzero determinant at all points of $\mathbf{A}$. If $f: g(\mathbf{A}) \to \mathbf{R}$ is an integrable function, then

$$\int_{g(\mathbf{A})} f = \int_{\mathbf{A}} (f \circ g)|\det g'|. \tag{4.24}$$

**Proof.** See [Spiv65] or [Buck78].

**4.8.7. Example.** Consider the map $g: \mathbf{R}^2 \to \mathbf{R}^2$ defined by

$$x = r\cos\theta$$
$$y = r\sin\theta$$

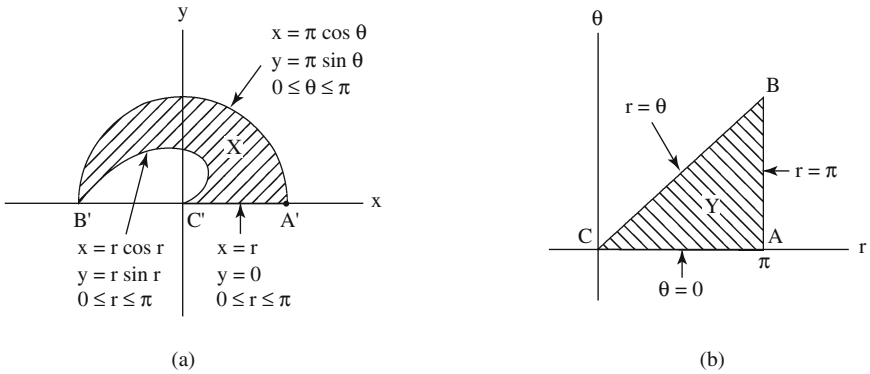



**Figure 4.25.** Change of variable Example 4.8.7.

which transforms from polar coordinates to Cartesian coordinates. The Jacobian matrix

$$g'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

has determinant r. Therefore, if $\mathbf{Y} \subset \mathbf{R}^2$ is a region in polar coordinate space and $\mathbf{X} = g(\mathbf{Y})$ is the corresponding region in Cartesian space, then equation (4.24) turns into

$$\int_{\mathbf{X}} f(x,y)dxdy = \int_{\mathbf{Y}} f(r\cos\theta, r\sin\theta)rdrd\theta.$$

As a special case, consider the region **Y** shown in Figure 4.25(b), which is bounded by the lines $r = \pi$, $r = \theta$, and $\theta = 0$. This region is mapped by g into the region **X** shown in Figure 4.25(a). If we were asked to compute an integral of a function f defined on the complicated region **X**, then, using our change of coordinate map g, we can translate the problem into computing a simple iterated integral over the region **Y**, that is,

$$\int_{\mathbf{X}} f(x,y)dxdy = \int_0^{\pi} r\left(\int_0^{r} f(r\cos\theta, r\sin\theta)d\theta\right)dr.$$

Theorem 4.8.6 proves two well-known facts. It demonstrates the geometric meaning of the determinant of a linear transformation as the factor by which volumes are changed. It also proves a formula for computing the volume of a parallelotope in $\mathbf{R}^n$ (Corollary 4.8.9 below).

**4.8.8. Corollary.** Let $T: \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation. If $\mathbf{A} \subset \mathbf{R}^n$ is an open bounded subset, then

$$\mathrm{vol}(T(\mathbf{A})) = |\det(T)|\mathrm{vol}(\mathbf{A}).$$

**Proof.** The definition of volume, Theorem 4.8.6, and the fact that the derivative of T is constant imply that

$$\mathrm{vol}(T(\mathbf{A})) = \int_{T(\mathbf{A})} 1_{T(\mathbf{A})} = \int_{\mathbf{A}} |\det T'| = |\det T'| \int_{\mathbf{A}} 1_{\mathbf{A}} = |\det(T)|\mathrm{vol}(\mathbf{A}).$$

**Note.** Some, but not all, proofs of Theorem 4.8.6, like the one given by Spivak ([Spiv65]), rely on having proved Corollary 4.8.8 as a special case when **A** is an open rectangle, so that the corollary would have to be proved separately. This is not hard to do. One can prove Corollary 4.8.8 by choosing some very simple linear transformations for which the result is trivial to prove and which have the property that any linear transformation is a composite of them. See [Spiv65].

**Definition.** Let $\mathbf{p}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbf{R}^n$. Assume that the vectors $\mathbf{v}_i$ are linearly independent. The set **X** in $\mathbf{R}^n$ defined by

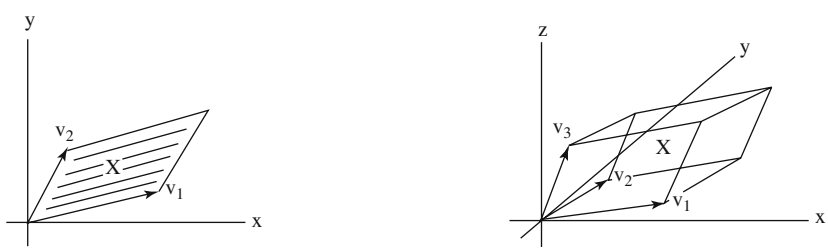
$$\mathbf{X} = \left\{ \mathbf{p} + \sum_{i=1}^{k} a_i \mathbf{v}_i \;\middle|\; 0 \le a_i \le 1 \right\}$$

is called a *parallelotope* or *parallelopiped* based at **p** and spanned by the $\mathbf{v}_i$. If the reference to **p** is omitted, then it is assumed that **p** is **0**. If k = 2, then **X** is also called a *parallelogram*. See Figure 4.26.

**4.8.9. Corollary.** Let **X** be a parallelotope in $\mathbf{R}^n$ based at **p** and spanned by some vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Then

$$\mathrm{vol}(\mathbf{X}) = \left| \det \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{pmatrix} \right|.$$

**Proof.** Because translation does not change volume, we may assume that **p** = **0**. Define a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ by $T(\mathbf{e}_i) = \mathbf{v}_i$. Clearly, T maps the unit cube, that is, the parallelotope based at **0** and spanned by $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, to **X**. Since the volume of the interior of the parallelotope is the same as the volume of the parallelotope, the result now follows from Corollary 4.8.8.




(a) parallelogram in $R^2$ (b) parallelotope in $R^3$

**Figure 4.26.** Parallelotopes **X** at the origin.

## 4.9 Differential Forms

The object of this section, the next section, and Section 8.12 is to try and make some sense out of differential symbols, such as "dx" and "dxdy", and to formalize the rules that the algebra of these differentials satisfies.

Most likely, the reader encountered such symbols very early in his or her calculus course. A common place is in integrals like

$$\int_a^b f(x)dx.$$

What really is the meaning of "dx"? Of course, in the integral above it probably was just pure notation. One could equally well just have written

$$\int_a^b f.$$

The typical reason for including the "dx" is to enable one to apply the chain rule more easily by reducing it to a formal symbol manipulating process: Given $x = g(u)$ one substitutes $g(u)$ for x and $g'(u)du$ for dx.

In two dimensions, things get more complicated. One often writes double integrals as

$$\iint_{\mathbf{X}} f(x,y)dxdy.$$

But just like the one-dimensional integrals, orientation plays a role and

$$\iint_{\mathbf{X}} f(x,y)dydx = -\iint_{\mathbf{X}} f(x,y)dxdy.$$

In terms of manipulating differentials, one can express this algebraically by saying that $dydx = -dxdy$.

Another place where one encounters the "dx" notation is in expressions such as

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

for functions f(x,y). This again has its uses, like the classical notation dy/dx for the derivative that is handy for changes of variables, but there come times, as the author himself found out on occasions as a student, when things get more complicated and such simple-minded notation involving differentials can get confusing. One ends up making some transformations that one has seen someone else make and that may be correct if one is lucky, but which one does not really understand. It therefore is useful to make sense of symbols such as "dx" in a rigorous way. Doing this will also help prevent mistakes in imprecise manipulations of these symbols. The "casual user" of differentials will gain nothing from this section and the next and probably should skip them. At the end we shall simply have justified the classic notation and manipulation

of differentials. However, by having expressed everything carefully, the reader who wants to use the notation in a serious way is less likely to make mistakes with it.

A limited and specialized development of differential forms for dimensions 1, 2, and 3 can be found in various advanced calculus books like [Buck78]. Such developments do not show the complete picture, however, and are more along the lines of applications or examples. A thorough discussion of differential forms is based on the "exterior algebra" of differential forms. Unfortunately, this involves a fair amount of abstract algebra, especially if one starts with an axiomatic approach to tensor and exterior algebras. We outline this approach in Appendix C. No matter how one develops the subject, the reader is forced to "suffer" through a large number of definitions and theorems. Although most of the theorems follow trivially from the definitions, it takes a while before one gets to the applications. The subject is actually a good example where the right mathematical definitions lead in the end to significant consequences, consequences that, because of the definitions, are trivial to prove. Many books on differentiable manifolds have a section on differential forms but most readers would probably find their presentations hard reading. In the opinion of the author, good references for differential forms are [Flan63], [Spiv65], and [Spiv70a]. Flanders does not prove everything but covers the subject and a great many applications well. He starts with an axiomatic formulation of the properties that one wants the exterior algebra to posses. The advantage to his approach is that one quickly starts using the usual notation associated with differential forms. Spivak has more limited goals but **does** prove all his results and does a good job in presenting the subject in a completely self-contained manner. He starts by defining the "tensor algebra" of multilinear maps and derives the exterior algebra from this. This more computational approach avoids the existence proofs and requires less familiarity with advanced algebraic concepts, but still forces a reader to wade through quite a few definitions and proofs of simple consequences before one gets to the actual differential forms themselves. By and large, our approach here will follow Spivak's, except that some of the algebraic preliminaries have been off-loaded to Appendix C. Another book worth looking at is [GuiP74]. This book also follows Spivak's approach. The reader who is overwhelmed by the many definitions and abstract concepts should look ahead to the end of this section and Section 8.12 to see that it will all be worth it.

Differential forms are derived from the exterior or Grassmann algebra $E(\mathbf{R}^{n*})$ of the dual space $\mathbf{R}^{n*}$. This algebra can be identified with the algebra $\Lambda(\mathbf{R}^n)$ of exterior forms on $\mathbf{R}^n$. The theoretical basis of these algebras is developed in Section C.6 and will not be reproduced here, but we summarize the essential aspects of the latter. The reader who is not interested in the mathematical details that justify our assertions can simply take them as axioms that specify a language of forms and how the symbols in that language are manipulated. All future properties will be deduced from these.

Let $\mathbf{V}$ be an n-dimensional vector space over the reals and $\mathbf{V}^*$ its dual.

**Notation.** Let $k \geq 0$. The vector space of alternating multilinear maps (also called *exterior k-forms of* $\mathbf{V}$)

$$\omega\colon \mathbf{V}^k = \underbrace{\mathbf{V} \times \mathbf{V} \times \ldots \times \mathbf{V}}_{k} \to \mathbf{R}.$$

is denoted by $\Lambda^k(\mathbf{V})$. By convention, $\mathbf{V}^0 = \mathbf{0}$ and $\Lambda^0(\mathbf{V}) = \mathbf{R}$. There is an *exterior* or *wedge product*

$$\wedge: \Lambda^r(\mathbf{V}) \times \Lambda^s(\mathbf{V}) \to \Lambda^{r+s}(\mathbf{V})$$

that is distributive and associative. (The classical notation such as dxdy is really an abbreviation of dx∧dy.) Define

$$\Lambda(\mathbf{V}) = \sum_{k=0}^{n} \Lambda^k(\mathbf{V}). \tag{4.25a}$$

Vector addition and the product $\wedge$ make $\Lambda(\mathbf{V})$ into an algebra called the *algebra of exterior forms* on $\mathbf{V}$,

**The defining properties of the algebra of exterior forms on V:**

(1) If $\omega \in \Lambda^r(\mathbf{V})$ and $\eta \in \Lambda^s(\mathbf{V})$, then

$$\omega \wedge \eta = (-1)^{rs} \eta \wedge \omega. \tag{4.25b}$$

In fact, for any permutation $\sigma$ of $\{1,2, \ldots ,k\}$ and $\alpha_i \in \Lambda^1(\mathbf{V})$, we have that

$$\alpha_{\sigma(1)} \wedge \alpha_{\sigma(2)} \wedge \ldots \wedge \alpha_{\sigma(k)} = (\text{sign } \sigma)\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_k. \tag{4.25c}$$

(2) Let $\alpha_1, \alpha_2, \ldots,$ and $\alpha_n$ form a basis for $\mathbf{V}^*$. If $1 \le k \le n$, then the set of all

$$\alpha_{i_1} \wedge \alpha_{i_2} \wedge \ldots \wedge \alpha_{i_k}, 1 \le i_1 < i_2 < \ldots < i_k \le n,$$

is a basis for $\Lambda^k(\mathbf{V})$. It follows that $\Lambda^k(\mathbf{V})$ has dimension $\binom{n}{k}$.

(3) If $\alpha_i \in \mathbf{V}^*$, then the element $\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_k \in \Lambda^k(\mathbf{V})$ satisfies

$$(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_k)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \det(\alpha_i(\mathbf{v}_j)) \tag{4.25d}$$

for all $\mathbf{v}_j \in \mathbf{V}$.

(4) If $T: \mathbf{V} \to \mathbf{W}$ is a linear transformation, then the induced map $T^*: \mathbf{W}^* \to \mathbf{V}^*$ on dual spaces defines an *induced map*

$$T^*: \Lambda^k(\mathbf{W}) \to \Lambda^k(\mathbf{V}), \tag{4.25e}$$

satisfying

$$T^*(\alpha)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \alpha(T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_k)), \alpha \in \Lambda^k(\mathbf{W}), \mathbf{v}_i \in \mathbf{V} \tag{4.25f}$$

$$T^*(\alpha \wedge \beta) = T^*(\alpha) \wedge T^*(\beta), \alpha \in \Lambda^r(\mathbf{W}), \beta \in \Lambda^s(\mathbf{W}). \tag{4.25g}$$

We summarize a few basic consequences of the above for emphasis.

**4.9.1. Proposition.**

(1) $\Lambda^0(\mathbf{V}) = \mathbf{R}$ and $\Lambda^1(\mathbf{V}) = \mathbf{V}^*$.
(2) If $\alpha_1, \alpha_2, \ldots,$ and $\alpha_n$ are a basis for $\mathbf{V}^*$, then the element $\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n$ is a basis for $\Lambda^n(\mathbf{V})$. In particular,

$$\Lambda^n(\mathbf{V}) \approx \mathbf{R}.$$

(3) The determinant map

$$\det: \mathbf{R}^n \to \mathbf{R}$$

is a basis for $\Lambda^n(\mathbf{R}^n)$. More generally, the maps

$$(\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k) \to \text{determinant of some fixed } k \times k \text{ minor of } \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}, \mathbf{v}_i \in \mathbf{R}^n,$$

are a basis for $\Lambda^k(\mathbf{R}^n)$.
(4) $\Lambda^k(\mathbf{V}) = \mathbf{0}$, for $k > n$.
(5) $\alpha \wedge \alpha = 0$ for all $\alpha \in \Lambda^k(\mathbf{V})$. If $\alpha, \beta \in \Lambda^1(\mathbf{V})$, then $\alpha \wedge \beta = -\beta \wedge \alpha$. In particular, the product $\wedge$ is not commutative.

**Proof.** The listed facts are easy consequences of the definitions and Properties (1)–(3) of $\wedge$. We shall only prove fact (3). Specializing fact (2) to $\mathbf{R}^n$ implies that if we pick any basis $\alpha_1, \alpha_2, \ldots,$ and $\alpha_n$ for $\mathbf{R}^{n*}$, then

$$\det = c\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n,$$

for some $c \in \mathbf{R}$. Now if we choose $\alpha_1, \alpha_2, \ldots,$ and $\alpha_n$ to be the dual basis for a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots,$ and $\mathbf{v}_n$ for $\mathbf{R}^n$, then Property (3) of $\wedge$ implies that

$$(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 1.$$

Choosing $\mathbf{v}_i = \mathbf{e}_i$, shows that

$$\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = c(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n)(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = c \cdot 1.$$

In other words, $c = 1$ and the first part of fact (3) is proved. The second part is proved in a similar fashion.

Proposition 4.9.1(4) explains why the summation in (4.25a) stops at n, since all the other spaces are $\mathbf{0}$.

Returning to $\mathbf{R}^n$, it is convenient to think of $\Lambda(\mathbf{R}^n)$ as the exterior form algebra associated to the origin of $\mathbf{R}^n$, because the coefficients of the elements are just reals. We would like to extend the exterior algebra notion by allowing functions $f: \mathbf{R}^n \to \mathbf{R}$

to be the coefficients in the sense that we would be specifying a collection of elements of $\Lambda(\mathbf{R}^n)$, one for each point of $\mathbf{R}^n$. The element that we get at any point in $\mathbf{R}^n$ is the one where we replace the function coefficient with the real number that is the value of the function at that point. This would be an adequate approach if one was only interested in defining differential forms for $\mathbf{R}^n$, but because we have the generalization to manifolds in mind (Section 8.12) we shall take a different approach. Rather than having global functions induce "local" elements of $\Lambda(\mathbf{R}^n)$, we shall build up a global structure from local ones.

For $\mathbf{p} \in \mathbf{R}^n$ define

$$T_{\mathbf{p}}(\mathbf{R}^n) = \{(\mathbf{p}, \mathbf{v}) \mid \mathbf{v} \in \mathbf{R}^n\}. \tag{4.26}$$

We endow the set $T_{\mathbf{p}}(\mathbf{R}^n)$ with the natural vector space structure defined by

$$(\mathbf{p}, \mathbf{v}) + (\mathbf{p}, \mathbf{v}') = (\mathbf{p}, \mathbf{v} + \mathbf{v}'), \mathbf{v}, \mathbf{v}' \in \mathbf{R}^n,$$

and

$$c(\mathbf{p}, \mathbf{v}) = (\mathbf{p}, c\mathbf{v}), c \in \mathbf{R}, \mathbf{v} \in \mathbf{R}^n.$$

**Definition.** The vector space $T_{\mathbf{p}}(\mathbf{R}^n)$ is called the *tangent space* to $\mathbf{R}^n$ at the point $\mathbf{p}$. The vector $(\mathbf{p},\mathbf{v})$ in $T_{\mathbf{p}}(\mathbf{R}^n)$, often denoted by $\mathbf{v}_{\mathbf{p}}$, is called a *tangent vector* to $\mathbf{R}^n$ at $\mathbf{p}$.

The space $T_{\mathbf{p}}(\mathbf{R}^n)$ is simply a way of formalizing the notion of each point of $\mathbf{R}^n$ having its **own** copy of $\mathbf{R}^n$ thought of as the tangent vectors at that point. One can think of the vector $\mathbf{v}_{\mathbf{p}}$ as a vector "starting" at $\mathbf{p}$. See Figure 4.27. In that sense, the tangent spaces also formalize the way one usually deals with vectors as arrows that seem to float around in space. We shall have lots more to say about tangent vectors and tangent spaces in the more general setting of curves, surfaces, and manifolds in Chapter 8.

Differentiable maps induce maps on tangent spaces.

**Definition.** Let $f: \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable function. For each $\mathbf{p} \in \mathbf{R}^n$, the derivative of f defines a linear transformation $Df(\mathbf{p}): \mathbf{R}^n \to \mathbf{R}^m$. The linear transformation

$$f_*: T_{\mathbf{p}}(\mathbf{R}^n) \to T_{f(\mathbf{p})}(\mathbf{R}^m)$$

defined by

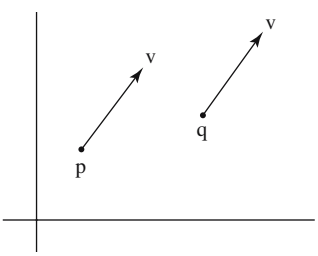



**Figure 4.27.** Vectors starting at different points.

$$f_*(\mathbf{v_p}) = (Df(\mathbf{p})(\mathbf{v}))_{f(\mathbf{p})}.$$

is called the *induced map* on tangent spaces.

**Definition.** A *vector field* on $\mathbf{R}^n$ is a map F defined on $\mathbf{R}^n$ that sends $\mathbf{p} \in \mathbf{R}^n$ to an element of $T_{\mathbf{p}}(\mathbf{R}^n)$. If we express F in the form

$$F(\mathbf{p}) = \sum_{i=1}^{n} F_i(\mathbf{p})(\mathbf{e}_i)_{\mathbf{p}}, \tag{4.27}$$

then the real-valued functions $F_i(\mathbf{p})$ are called the *component functions* of F.

**Definition.** A *differential k-form* on $\mathbf{R}^n$, or simply *k-form* or *differential form*, is a map $\omega$ defined on $\mathbf{R}^n$ that sends $\mathbf{p} \in \mathbf{R}^n$ to an element $\omega(\mathbf{p}) \in \Lambda^k(T_{\mathbf{p}}(\mathbf{R}^n))$.

The set of k-forms on $\mathbf{R}^n$ actually forms a vector space if we define the addition and scalar multiplication in a pointwise fashion. Note that differential 0-forms are just functions $\mathbf{R}^n \to \mathbf{R}$. Also, given a differentiable function $f: \mathbf{R}^n \to \mathbf{R}$, $Df(\mathbf{p})$ is linear transformation from $\mathbf{R}^n$ to $\mathbf{R}$. As such it can be considered an element of $\Lambda^1(\mathbf{R}^n)$, in fact, an element of $\Lambda^1(T_{\mathbf{p}}(\mathbf{R}^n))$.

**Definition.** The *differential* of a function $f: \mathbf{R}^n \to \mathbf{R}$, denoted by df, is the differential 1-form in $\Lambda^1(\mathbf{R}^n)$ defined by

$$df(\mathbf{p})(\mathbf{v_p}) = Df(\mathbf{p})(\mathbf{v}). \tag{4.28}$$

As a special case, consider the projection functions $\pi_i: \mathbf{R}^n \to \mathbf{R}$ defined by $\pi_i(x_1, x_2, \ldots, x_n) = x_i$.

**Notation.** In order to arrive at the classical notation for differential forms we shall abuse the notation and write $dx_i$ instead of $d\pi_i$.

Now, if $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, then

$$dx_i(\mathbf{p})(\mathbf{v_p}) = Dx_i(\mathbf{p})(\mathbf{v}) = v_i.$$

Therefore, the linear maps $dx_i(\mathbf{p})$ are just the dual basis of the standard basis $(\mathbf{e}_1)_{\mathbf{p}}$, $(\mathbf{e}_2)_{\mathbf{p}}, \ldots, (\mathbf{e}_n)_{\mathbf{p}}$ of $T_{\mathbf{p}}(\mathbf{R}^n)$, so that every differential k-form $\omega$ can be expressed in the form

$$\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} \omega_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}, \tag{4.29}$$

for functions $\omega_{i_1 \ldots i_k}: \mathbf{R}^n \to \mathbf{R}$. In particular, every n-form $\omega$ has the form

$$\omega = f\, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n, \tag{4.30}$$

for some function $f: \mathbf{R}^n \to \mathbf{R}$.

**Definition.** The differential form $\omega$ is called *continuous*, *differentiable*, $C^\infty$, etc., if the functions $\omega_{i_1 \ldots i_k}$ are continuous, differentiable, $C^\infty$, etc., respectively.

We shall always assume that differentiable forms are $C^\infty$ in order to avoid problems with functions not being differentiable enough.

**Notation.** The vector space of $C^\infty$ differential k-forms on $\mathbf{R}^n$ will be denoted by $\Omega^k(\mathbf{R}^n)$. (We do not use the notation $\Lambda^k(\mathbf{R}^n)$ because that already refers to the alternating multilinear maps of the **vector space** $\mathbf{R}^n$, which is something quite different.)

The next theorem expands on the result expressed by equation (4.29) by describing the expansion in more detail for the case of differentials of functions.

**4.9.2. Theorem.** If $f: \mathbf{R}^n \to \mathbf{R}$ is a differentiable function, then

$$df = D_1 f\, dx_1 + D_2 f\, dx_2 + \ldots + D_n f\, dx_n,$$

that is, using the classical notation,

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \ldots + \frac{\partial f}{\partial x_n} dx_n. \tag{4.31a}$$

In particular, the differential df can be expressed in terms of the directional derivative by the formula

$$df(\mathbf{p})(\mathbf{v_p}) = D_{\mathbf{v}} f(\mathbf{p}). \tag{4.31b}$$

**Proof.** Simply note that

$$df(\mathbf{p})(\mathbf{v_p}) = Df(\mathbf{p})(\mathbf{v}) = \sum_{i=1}^{n} v_i D_i f(\mathbf{p}) = \sum_{i=1}^{n} dx_i(\mathbf{p})(\mathbf{v_p}) D_i f(\mathbf{p})$$

and

$$\sum_{i=1}^{n} v_i D_i f(\mathbf{p}) = \nabla f(\mathbf{p}) \bullet \mathbf{v} = D_{\mathbf{v}} f(\mathbf{p}).$$

Next, consider a differentiable map $f: \mathbf{R}^n \to \mathbf{R}^m$. The map f induces a map $f_*$ on tangent spaces and its dual map

$$f^*: T_{f(\mathbf{p})}(\mathbf{R}^m)^* \to T_{\mathbf{p}}(\mathbf{R}^n)^*$$

leads to the induced map

$$f^*: \Lambda^k\left(T_{f(\mathbf{p})}(\mathbf{R}^m)\right) \to \Lambda^k\left(T_{\mathbf{p}}(\mathbf{R}^n)\right). \tag{4.32}$$

**Definition.** The map

$$f^*: \Omega^k(\mathbf{R}^m) \to \Omega^k(\mathbf{R}^n)$$

defined by

$$(f^*\omega)(\mathbf{p}) = f^*(\omega(f(\mathbf{p})), \tag{4.33}$$

is called the *induced map* on differential forms. (The map $f^*$ on the right-hand side of equation (4.33) is the induced map referred to by expression (4.32).) The form $f^*\omega$ is often called the *pullback* to $\mathbf{R}^n$ of the form $\omega$ on $\mathbf{R}^m$.

Using equation (4.25f), the action of the induced map $f^*$ on differential k-forms can be described in more detail as follows: Let $\mathbf{v}_i \in \mathbf{R}^n$. Then

$$(f^*\omega)(\mathbf{p})\big((\mathbf{v}_1)_\mathbf{p}, (\mathbf{v}_2)_\mathbf{p}, \ldots, (\mathbf{v}_k)_\mathbf{p}\big) = \omega\big(f(\mathbf{p})\big(f_*\big((\mathbf{v}_1)_\mathbf{p}\big), f_*\big((\mathbf{v}_2)_\mathbf{p}\big), \ldots, f_*\big((\mathbf{v}_k)_\mathbf{p}\big)\big). \tag{4.34}$$

The next theorem lists the main properties of the induced map $f^*$ that enables us to compute the map easily.

**4.9.3. Theorem.** If $f: \mathbf{R}^n \to \mathbf{R}^m$ is a differentiable function, then the induced map

$$f^*: \Omega^k(\mathbf{R}^m) \to \Omega^k(\mathbf{R}^n)$$

on differentiable forms satisfies

(1) $f^*(\omega_1 + \omega_2) = f^*(\omega_1) + f^*(\omega_2)$
(2) $f^*(g\omega) = (g \circ f)f^*\omega$
(3) $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$
(4) $f^*(dx_i) = \sum_{j=1}^n D_j f_i \, dx_j = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} dx_j$
(5) $f^*(g \, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n) = (g \circ f)(\det f') \, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n$

**Proof.** The proofs of (1)–(4) are simply a case of expanding all the expressions using their definitions. As an example, (4) is proved by the following equalities:

$$\begin{aligned}
f^*(dx_i)(\mathbf{p})(\mathbf{v}_\mathbf{p}) &= dx_i(f(\mathbf{p}))(f_*(\mathbf{v}_\mathbf{p})) \\
&= dx_i(f(\mathbf{p}))\left(\sum_{j=1}^n v_j \, D_j f_1(\mathbf{p}), \sum_{j=1}^n v_j \, D_j f_2(\mathbf{p}), \ldots, \sum_{j=1}^n v_j \, D_j f_m(\mathbf{p})\right) \\
&= \sum_{j=1}^n v_j \, D_j f_i(\mathbf{p}) \\
&= \sum_{j=1}^n D_j f_i(\mathbf{p}) \, dx_j(\mathbf{p})(\mathbf{v}_\mathbf{p}).
\end{aligned}$$

For a proof of (5), see [Spiv65].

Next, we generalize the differential of a function as defined by equation (4.31a) to a differential of an arbitrary differential form.

**Definition.** Given a k-form

$$\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} \omega_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}$$

define a (k + 1)-form $d\omega$, called the *differential* of $\omega$, by

$$\begin{aligned} d\omega &= \sum_{1 \le i_1 < \ldots < i_k \le n} d\omega_{i_1 \ldots i_k} \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k} \\ &= \sum_{1 \le i_1 < \ldots < i_k \le n} \sum_{j=1}^{n} D_j(\omega_{i_1 \ldots i_k})\, dx_j \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}. \end{aligned}$$

The map $\omega \to d\omega$ is called the *differential operator* d for differential forms.

Note that if we consider a function $f : \mathbf{R}^n \to \mathbf{R}$ as a 0-form, then the new definition of the differential of f agrees with our earlier one.

**4.9.4. Theorem.**

(1) If $\omega$ and $\eta$ are two k-forms, then $d(\omega + \eta) = d\omega + d\eta$.
(2) If $\omega$ is an r-form and $\eta$ is an s-form, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^{rs} \omega \wedge d\eta.$$

(3) $d(d\omega) = 0$
(4) If $f : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable and $\omega$ is a k-form, then $f^*(d\omega) = d(f^*\omega)$.

**Proof.** For fact (2), check the formula first on the 1-forms $dx_i$ and their wedge products. Fact (3) is proved by direct computation using formula (4.25b) that will cause terms to cancel. Fact (4) is proved by induction on k. See [Spiv65].

One can show that Theorem 4.9.4(1)–(3) and equation (4.31a) can be considered axioms for the differential operator d that define it uniquely.

We now know all the basic facts we need to know about differential forms and are ready to move on to the important integration applications. Before we do though, we finish this section with definitions of several well-known concepts.

**Definition.** Let F be a vector field on $\mathbf{R}^n$ with component functions $F_i$. The *divergence* of F, denoted by div F, is defined by

$$\operatorname{div} F = \sum_{i=1}^{n} D_i F_i.$$

If n = 3, then the *curl* of F, denoted by curl F, is the vector field on $\mathbf{R}^3$ defined by

$$(\text{curl } F)(\mathbf{p}) = (D_2F_3 - D_3F_2)(\mathbf{e}_1)_{\mathbf{p}} + (D_3F_1 - D_1F_3)(\mathbf{e}_2)_{\mathbf{p}} + (D_1F_2 - D_2F_1)(\mathbf{e}_3)_{\mathbf{p}}.$$

In classical notation,

$$\text{div } F = \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i} = \nabla \bullet F$$
$$\text{curl } F = \nabla \times F.$$

The comment earlier and Exercises 4.9.4 and 4.9.5 basically show that the gradient, curl, and divergence operators are just the differential operator d.

### 4.9.1 Differential Forms and Integration

The last section developed the basic framework of differential forms. In this section, and then later in Section 8.12, we show how differential forms are related to integration. Here we shall deal with integration on open subsets of $\mathbf{R}^n$. Section 8.12 will deal with integration on manifolds. We will have to plow through a lot of definitions and technical facts, but at the end it will all be worth it because some important theorems like Stokes' theorem will become trivialities! We shall follow the presentation given in [Spiv65] and [Spiv70a].

Although all definitions and results were stated with respect to $\mathbf{R}^n$ in the last section, this was done only to simplify the discussion. It is easy to check that everything applies to open subsets. We shall need this more general context now. In fact, we shall also feel free to talk about differential forms on cubes $[0,1]^n$. Although the latter are closed sets, they are very nice sets and when it comes to differentiability issues one can either take one-sided derivatives or assume without any problem that functions are defined on an open neighborhood of the cube.

Let $\mathbf{A}$ be an open subset of $\mathbf{R}^n$.

**Definition.** A *singular k-cube* in $\mathbf{A}$ is a $C^\infty$ function $c:[0,1]^k \to \mathbf{A}$.

As usual, we chose singular k-cubes to be $C^\infty$ so that we do not have to mess with differentiability questions. A singular 0-cube can be thought of as a point and a singular 1-cube is just a parametric curve. We shall need to talk about boundaries of cubes and so we need some more notation.

**Definition.** The inclusion map $I^n:[0,1]^n \to \mathbf{R}^n$, $I^n(\mathbf{p}) = \mathbf{p}$, is called the *standard singular n-cube*. The *(i,j)-faces* of $I^n$ are the singular $(n-1)$-cubes

$$I^n_{(i,j)}:[0,1]^{n-1} \to [0,1]^n,\ 1 \le i \le n \text{ and } j = 0 \text{ or } 1,$$

defined by

$$I^n_{(i,j)}(x_1, x_2, \ldots, x_{n-1}) = (x_1, x_2, \ldots, x_{i-1}, j, x_i, \ldots, x_{n-1}). \tag{4.35}$$

Given a singular k-cube $c:[0,1]^k \to \mathbf{A}$, define the singular $(n-1)$-cube $c_{i,j}$, called the *(i,j)-face* of c, by

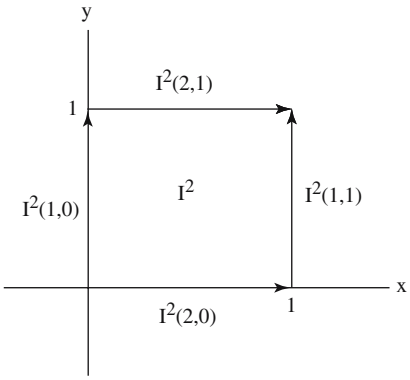



**Figure 4.28.** The faces of $I^2$.

$$c_{i,j} = c \circ I^k_{(i,j)}.$$

See Figure 4.28 for the (i,0)- and (i,1)-faces of $I^2$. The arrows on the edges show the orientations that the functions $I^2_{(i,j)}$ induce on the edges.

**Definition.** A formal linear combinations of singular k-cubes for an open set **A** is called a *singular k-chain* and the set of these is denoted by $\Gamma_k(\mathbf{A})$.

We do not give a precise definition of "formal linear combinations" here. We basically want to write formal expressions such as $2c_1 - 3c_2 + c_3$, where $c_1$, $c_2$, and $c_3$ are singular k-cubes. The interested reader can look ahead to Section 7.2.1 where we give a precise definition for a similar concept in the context of chain groups $C_k(K)$ for a simplicial complex K.

**Definition.** Given a singular k-cube $c:[0,1]^k \to \mathbf{A}$, define the singular $(k-1)$-chain $\partial c$, called the *boundary of c*, by

$$\partial c = \sum_{i=1}^{k} \sum_{j=0}^{1} (-1)^{i+j} c_{i,j}. \tag{4.36}$$

More generally, define the *boundary operator*

$$\partial : \Gamma_k(\mathbf{A}) \to \Gamma_{k-1}(\mathbf{A})$$

by

$$\partial\left(\sum_i a_i c_i\right) = \sum_i a_i \partial c_i. \tag{4.37}$$

The signs associated to the faces in the boundary expression (4.36) of a cube should be interpreted as indicating the orientation that the cube induces on the face.

Consider Figure 4.28 again. The standard orientation of $I^2$ induces a counterclockwise orientation on the edges. Formula (4.36) implies that

$$\partial I^2 = -I^2_{(1,0)} + I^2_{(1,1)} + I^2_{(2,0)} - I^2_{(2,1)},$$

which specifies that same orientation.

**4.9.1.1. Lemma.** $\partial \circ \partial = 0$.

**Proof.** This is a straightforward computation.

Lemma 4.9.1.1 is only an aside remark for us here, but it actually is the foundation of an important theory. To do this theory justice would take us far afield and take too much time. Our objectives in this section are much more limited. Nevertheless, the reader should return here after reading Chapter 7. We are basically describing a homology theory for open sets **A** similar to the homology theory for simplicial complexes that will be developed in Chapter 7. The main difference is that we are using singular k-cubes rather than k-simplices.

With these preliminaries out of the way, we return to the question of integration. Suppose that $\omega$ is a k-form on $[0,1]^k$. In this case we know that

$$\omega = f\, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_k,$$

for some unique function $f:[0,1]^k \to \mathbf{R}$.

**Definition.** Define the *integral of $\omega$ over $[0,1]^k$* by

$$\int_{[0,1]^k} \omega = \int_{[0,1]^k} f. \tag{4.38}$$

Of course, the integral on the right-hand side of equation (4.38) is the standard advanced calculus integral that was defined in Section 4.8. Note how similar all of the notation is, that is, we are saying that

$$\int_{[0,1]^k} f\, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_k = \int_{[0,1]^k} f\, dx_1 dx_2 \ldots dx_k.$$

Finally,

**Definition.** Let $k > 0$. If $\omega$ is a k-form on **A** and if $c:[0,1]^k \to \mathbf{A}$ is a singular k-cube, then define the *integral of $\omega$ over $c$* by

$$\int_c \omega = \int_{[0,1]^k} c^*\omega. \tag{4.39a}$$

If $\omega$ is a 0-form on **A**, then $\omega$ is just a function $f:\mathbf{A} \to \mathbf{R}$. Therefore, if $c:0 \to \mathbf{A}$ is a singular 0-cube, then define

$$\int_c \omega = \omega(c(0)). \tag{4.39b}$$

The *integral of* $\omega$ *over singular k-chain*

$$c = \sum_i a_i c_i$$

is defined by

$$\int_c \omega = \sum_i a_i \int_{c_i} \omega. \tag{4.39c}$$

Again, if one writes out the expressions in detail they will not seem much dissimilar from those of advanced calculus.

**4.9.1.2. Example.** Integrals of singular 1-cubes are nothing but what are called line integrals. To see this, assume that $\mathbf{A} \subseteq \mathbf{R}^2$ and consider a 1-form $\omega$ on $\mathbf{A}$. We can write $\omega$ in the form

$$\omega = a\,dx + b\,dy,$$

for some functions $a, b : \mathbf{A} \to \mathbf{R}$. In classical language, the integral of $\omega$ along (over) a singular 1-cube (curve) $c : [0,1] \to \mathbf{A}$ is called a *line integral* along the curve c. Furthermore, if $c(t) = (c_1(t), c_2(t))$, then

$$\begin{aligned}\int_c \omega &= \int_{[0,1]} \omega(c'(t))dt = \int_0^1 \omega(c'(t))dt \\ &= \int_0^1 [a(c(t))c_1'(t) + b(c(t))c_2'(t)]dt.\end{aligned}$$

Line integrals are often used in physics. For example, one might want to integrate a force field along curve.

**4.9.1.3. Example.** Consider integrals of 2-cubes, which correspond to classical surface integrals. Assume that $\mathbf{A} \subseteq \mathbf{R}^3$ and consider a 2-form $\omega$ on $\mathbf{A}$. Let $c : [0,1]^2 \to \mathbf{A}$ be a singular 2-cube in $\mathbf{A}$.

$$\iint_c \omega = \iint_{I^2} c^*\omega = \int_0^1 \int_0^1 \omega(c_x, c_y)dxdy,$$

where

$$c_x = \frac{\partial c}{\partial x} \quad \text{and} \quad c_y = \frac{\partial c}{\partial y}.$$

We can write $\omega$ in the form

$$\omega = a\,dx \wedge dy + b\,dx \wedge dz + c\,dy \wedge dz,$$

but repeating what we did in Example 4.9.1.2 would now get very messy and so we shall not attempt to do so.

**4.9.1.4. Theorem.** (Stokes' Theorem) Let $\mathbf{A} \subseteq \mathbf{R}^n$. If $\omega$ is a $(k-1)$-form on $\mathbf{A}$ and if c is a k-chain on $\mathbf{A}$, then

$$\int_c d\omega = \int_{\partial c} \omega.$$

**Proof.** See [Spiv65]. The proof is really not very hard and boils down mainly to using the definitions of the various quantities that are involved and showing that both sides of the equation are the same, starting first with the special case $c = I^k$.

If one writes out what Theorem 4.9.1.4 says in the case of the 1-chain $I^1$ in $\mathbf{R}$, one will see that what one has is just the classical Fundamental Theorem of Calculus (Theorem D.1.3). (Actually, the proof of Theorem 4.9.1.4 uses that special case and assumes that it has been proved separately.) One should therefore not be surprised when Theorem 4.9.1.4 is often called the *Generalized Fundamental Theorem of Calculus*. To quote from [Spiv65]:

(1) It is trivial.
(2) It is trivial because the terms in it have been properly defined.
(3) It has significant consequences.

One sees the truth of point (2) over and over in mathematics. Making the right definitions can isolate the essential aspects needed to arrive at a solution to a problem. The validity of point (3) in our current context of integration will become more apparent in Section 8.12. For now we have finished with our outline of differential forms and their relation to integration.

## 4.10 Exercises

### Section 4.2

4.2.1. Prove that the sets $\mathbf{R}^n$ and $\phi$ are **both** open and closed.

4.2.2. Prove that a single point is always a closed set.

4.2.3. Is $\{1/n \mid n = 1, 2, \ldots\}$ a closed set? Prove or disprove your answer.

4.2.4. Prove that if $\mathbf{X}$ is a closed set in $\mathbf{R}^n$, then $\mathrm{cl}(\mathbf{X}) = \mathbf{X}$.

4.2.5. Prove that the interior of a set is an open set.

4.2.6. Give examples of sets that show the following statements are false:

(a) If $\mathbf{X} \subseteq \mathbf{Y}$, then $\mathrm{bdry}(\mathbf{X}) \subseteq \mathrm{bdry}(\mathbf{Y})$.
(b) $\mathrm{bdry}(\mathbf{X}) = \mathrm{bdry}(\mathrm{cl}(\mathbf{X}))$.
(c) $\mathrm{bdry}(\mathbf{X}) = \mathrm{bdry}(\mathrm{int}(\mathbf{X}))$.
(d) $\mathrm{int}(\mathbf{X}) = \mathrm{int}(\mathrm{cl}(\mathbf{X}))$.

4.2.7. Prove that (0,1] is open in [–1,1].

4.2.8. Prove that any set is open or closed in itself.

4.2.9. Prove that the function $f:(0,1] \to \mathbf{R}$, $f(x) = 1/x$, is continuous but not uniformly continuous.

4.2.10. Let $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{R}^2$. Define a homeomorphism between the simplex $\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2$ and the unit disk $\mathbf{D}^2$.

4.2.11. Show that the support of a function is the intersection of all closed sets $\mathbf{A}$, where f vanishes outside of $\mathbf{A}$.

## Section 4.3

4.3.1. Prove Proposition 4.3.4.

4.3.2. Prove Corollary 4.3.7(2).

4.3.3. Let $f(t) = (t^2,t)$, $g(x,y) = y^2 - 4x$, and $G(t) = g(f(t))$. Compute Df, Dg, and DG. Determine DG in two ways: from its formula and by using the chain rule.

4.3.4. Let $f(x,y) = (x^3 - 2x + 1,\ xy + y^2,\ x - 3y + 7)$. Compute the Jacobian matrix $f'(-1,5)$. What is its rank?

4.3.5. Let $A, B:[a,b] \to \mathbf{R}^3$ be differentiable functions and define $f:[a,b] \to \mathbf{R}^3$ by $f(t) = A(t) \times B(t)$. Prove that

$$f'(t) = A'(t) \times B(t) + A(t) \times B'(t).$$

(We are treating the $1 \times 3$ Jacobian matrices as vectors here.) In short hand, the differentiation rule for the cross-product is $(A \times B)' = A' \times B + A \times B'$.

4.3.6. Show that the function $z = f(xy)$ satisfies the equation

$$x\frac{\partial z}{\partial x} - y\frac{\partial z}{\partial y} = 0.$$

4.3.7. Show that the substitution $x = e^s$ and $y = e^t$ converts the equation

$$x^2\frac{\partial^2 u}{\partial x^2} + y^2\frac{\partial^2 u}{\partial y^2} + x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = 0$$

into

$$\frac{\partial^2 u}{\partial s^2} + \frac{\partial^2 u}{\partial t^2} = 0.$$

4.3.8. If $f(x,y,z) = x \sin z$ and $\mathbf{v} = (2,-1,3)$, find $D_{\mathbf{v}}f(0,1,-1)$.

4.3.9. Show that the directional derivative of $f(x,y) = y^2/x$ at any point of the ellipse $2x^2 + y^2 = 1$ in the direction of the normal to the ellipse at that point is zero.

**Section 4.4**

4.4.1. Consider the curve **C** defined by the equation $x + y^2 + \cos xy = 0$.

(a) Can **C** be parameterized by a function of the form $y = f(x)$ in a neighborhood of (0,0)?
(b) Can **C** be parameterized by a function of the form $x = g(y)$ in a neighborhood of (0,0)?

4.4.2. The point $\mathbf{p} = (1,2,1)$ lies on the set **X** defined by

$$\begin{aligned} xy - 4xz + yz &= 0 \\ xyz + x + z - 4 &= 0. \end{aligned}$$

Determine which of the variables can be solved for in terms of the other two at **p**.

**Section 4.5**

4.5.1. Discuss the nature of the critical points of the function $f(x,y) = 2x^4 + y^4 - x^2 - 2y^2$.

4.5.2. Consider the function $f(x,y) = x^2 + 2xy - 4x + 8y$. Find the maxima, minima, and saddle points of f in the rectangle bounded by the lines $x = -5$, $x = 1$, $y = 0$, and $y = 7$.

4.5.3. Find the extreme value of the function $f(x,y,z) = xyz$ subject to the constraints

$$\frac{1}{x} + \frac{1}{y} + \frac{1}{z} = c > 0$$

and $x, y, z > 0$.

4.5.4. (a) Prove Case 1 for Theorem 4.5.13.
(b) Fill in the details left out of the proof of Case 2 for Theorem 4.5.13.

**Section 4.8**

4.8.1. Let **X** be the region of the plane defined by

$$\mathbf{X} = \left\{(x,y) \,\middle|\, 4 \le x^2 + y^2 \le 16,\ \sqrt{3}y - x \ge 0,\ \sqrt{3}x - y \ge 0\right\}.$$

If $f(x,y) = x$, compute the integral $\int_{\mathbf{X}} f$ by a change of variables to polar coordinates.

**Section 4.9**

4.9.1. If $f(x,y) = xy + \sin xy$, $\mathbf{v} = (-2,1)$, and $\mathbf{p} = (2,3)$, find $df(\mathbf{p})(\mathbf{v_p})$.

4.9.2. Define $f: \mathbf{R}^2 \to \mathbf{R}$ by $f(x,y) = xy + 3y$. Find $f_*(\mathbf{v_p})$, where $f_*: T_{\mathbf{p}}(\mathbf{R}^2) \to T_{f(\mathbf{p})}(\mathbf{R})$, $\mathbf{v} = (-2,1)$, and $\mathbf{p} = (3,5)$.

4.9.3. (a) If $\omega = xy\,dx + y^2\,dy$, compute $d\omega$.
(b) If $\omega = xy\,dxdy$, compute $d\omega$.

4.9.4. If

$$\omega = a\,dx + b\,dy + c\,dy$$

is a 1-form on $\mathbf{R}^3$, show that

$$d\omega = \left(\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y}\right)dx \wedge dy + \left(\frac{\partial c}{\partial x} - \frac{\partial a}{\partial z}\right)dx \wedge dz + \left(\frac{\partial c}{\partial y} - \frac{\partial b}{\partial z}\right)dy \wedge dz.$$

4.9.5. If

$$\omega = F_1\,dy \wedge dz + F_2\,dz \wedge dx + F_3\,dx \wedge dy$$

is a 2-form on $\mathbf{R}^3$, show that

$$d\omega = (\operatorname{div} F)\,dx \wedge dy \wedge dz,$$

where $F = (F_1, F_2, F_3)$.

**Section 4.9.1**

4.9.1.1. Compute the line integral

$$\int_c xy\,dx + y\,dy$$

along the curve c when

(a) c is the part of the parabola $y = x^2$ from (0,0) to (1,1).
(b) c is the polygonal path with vertices (0,0), (1,0), (1,1), (0,1), and (0,0).

CHAPTER 5

# Point Set Topology

## 5.1 Introduction

In this chapter we introduce the basic concepts dealing with metric and topological spaces and their associated maps. We shall build on the special case of $\mathbf{R}^n$ as described in Section 4.2. The reader new to topology can always think in terms of Euclidean space and its subspaces. That is certainly where currently most of the applications are. On the other hand, abstract topological spaces are not just abstract nonsense and it is worthwhile to introduce them even though metric spaces would be adequate for geometric modeling. The fact is that the metric usually has little to do with anything. The key concept is that of open sets. It is the open sets that really define a topology and by studying topological spaces one strips away the unimportant elements and gets to what is essential for understanding intrinsic topological properties of spaces. It is inadequate to think of spaces purely in terms of specific imbeddings in $\mathbf{R}^n$, even though, for example, the average person probably only thinks of a circle as an object sitting in some background like a piece of paper. If we want to study our universe, we would not think of it as imbedded in another space. The circle and universe have intrinsic properties that do not depend on any particular imbedding. One needs to see beyond the imbeddings.

Our intent is to survey only the most important results from what is called *general* or *point set topology* – those that get used a lot in other contexts. There are actually no really difficult theorems in this chapter. Most follow from the definitions in a relatively straightforward way. Of course, the author realizes that the material here may well be new to many readers and quite different from what they may have seen before, so that even easy results may seem hard initially. There are quite a few books on point set topology, but the one that this author recommends most highly is Eisenberg's book [Eise74]. Most of the explicit references in this chapter will be to this book, however, in the case of references for omitted proofs one can find these in many other books, such as [Lips65].

Metric spaces are certainly the most important topological spaces and we start with those in Section 5.2. Section 5.3 defines and discusses general topological spaces. Section 5.4 describes some important standard operations that create new spaces

from old ones. If one had to list the two most used properties of topological spaces it would certainly be compactness and connectedness and we devote Sections 5.5 and 5.6 to those, respectively. The basic problem of topology is to classify spaces up to homeomorphism and to find invariants that can be used to distinguish homeomorphism classes. However, along with deformations of spaces, it is also useful to study deformations of mappings and we do this in Section 5.7 where we discuss homotopy. Section 5.8 describes conditions for the existence of certain continuous functions. In Section 5.9 we take another look at a very important space, $\mathbf{P}^n$, and discuss some of its topological properties.

Finally, point set topology is one of those fields where one encounters a great many definitions. The reader may start to feel overwhelmed by all these definitions at the first reading of this chapter. In some sense, the reader can "ignore" them until they become relevant in the context of specific results. The reader may also run into many of the terms elsewhere, and so this chapter will serve as a general reference for what they mean. Certainly, we had to present them here because they represent certain technical conditions without which theorems would be false. The reader who is learning about topology for the first time may wonder "what the fuss is all about" because the conditions might seem like conditions that should obviously hold. They may in fact hold for all the "nice" spaces we will ever consider. However, definitions by their nature are abstract and they may have consequences that were unintended. For example, the definition of a continuous function is one with which "everyone" is happy and which can be used very effectively, but there are continuous functions that are nowhere differentiable. Is that what one had in mind when defining a continuous function? (As an aside, although such functions are undesirable, their existence actually gives us insight into what continuity really means.) The same thing is true here. Point set topology is a very large field. The definitions that we shall deal with in this book only scratch the surface. The fact is that, like the continuity of functions, the definition of a topology on a space is abstract and although it captures a basic idea that was extrapolated from nice subspaces of $\mathbf{R}^n$, it allows for a much larger universe of spaces. This is why we shall have to introduce some additional conditions from time to time (like requiring the differentiability of a function in addition to its continuity) to guarantee that we get what we want. Isolating needed conditions for a result and giving them a name is helpful in understanding what makes a theorem true. Unfortunately, this book will often not have time to explain fully the nuances and reasons for naming special conditions leaving some readers with a feeling of mystery. The only cure for such feelings is to read one of the more comprehensive books on topology listed in the references.

## 5.2 Metric Spaces

**Definition.** A *metric* on a set $\mathbf{X}$ is a function

$$\mathrm{d}: \mathbf{X} \times \mathbf{X} \to \mathbf{R}$$

such that the following holds for all $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbf{X}$:

(1) $d(\mathbf{p},\mathbf{q}) \geq 0$.
(2) $d(\mathbf{p},\mathbf{q}) = 0$ if and only if $\mathbf{p} = \mathbf{q}$.
(3) (symmetry) $d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p})$.
(4) (triangle inequality) $d(\mathbf{p},\mathbf{r}) \leq d(\mathbf{p},\mathbf{q}) + d(\mathbf{q},\mathbf{r})$.

The value $d(\mathbf{p},\mathbf{q})$ is called the *d-distance from* $\boldsymbol{p}$ *to* $\boldsymbol{q}$. A *metric space* is a pair $(\mathbf{X},d)$, where d is a metric on $\mathbf{X}$.

$\mathbf{R}^n$ admits a number of different metrics. Let $\mathbf{p}, \mathbf{q} \in \mathbf{R}^n$.

The *standard Euclidean metric* d: $d(\mathbf{p},\mathbf{q}) = |\mathbf{pq}|$

The *taxicab metric* $d_1$: $$d_1(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} |p_i - q_i|$$

The *max metric* $d_\infty$: $$d_\infty(\mathbf{p},\mathbf{q}) = \max_{1\leq i\leq n}\{|p_i - q_i|\}$$

**Note.** Throughout this book, whenever we talk about $\mathbf{R}^n$, we shall always assume that its metric is the Euclidean metric unless it is explicitly stated otherwise. The definitions and concepts generalize those in Section 4.2.

Many other spaces have metrics. A large source of metric spaces are vector spaces with an inner product because the distance function between vectors defined by the inner product is one (Exercise 5.2.2). This applies in particular to many function spaces.

**5.2.1. Example.** The space $C^0([0,1])$ of continuous functions on [0,1] can be made into a vector space by defining the addition of functions and scalar multiplication in a pointwise fashion. It is easy to check (Exercise 5.2.3) that one possible inner product on this space is defined by the formula

$$<f,g> = \int_0^1 f(t)g(t)dt. \tag{5.1}$$

This inner product leads to the metric

$$d_2(f,g) = \left(\int_0^1 (f(t)-g(t))^2 dt\right)^{1/2}. \tag{5.2}$$

Two other metrics on $C^0([0,1])$ unrelated to the inner product are

$$d_1(f,g) = \int_0^1 |f(t)-g(t)|dt \tag{5.3}$$

and

$$d_\infty(f,g) = \max_{0\leq t\leq 1} |f(t)-g(t)|. \tag{5.4}$$

See Exercises 5.2.4 and 5.2.5.

**5.2.2. Example.** Let **X** be any set and define a map

$$\delta : \mathbf{X} \times \mathbf{X} \to \mathbf{R}$$

by

$$\begin{aligned} \delta(\mathbf{p},\mathbf{q}) &= 0, \quad \text{if} \quad \mathbf{p} = \mathbf{q} \\ &= 1, \quad \text{otherwise.} \end{aligned}$$

It is easy to show that $\delta$ is a metric on **X**.

**Definition.** The map $\delta$ is called the *discrete* metric on **X**.

The discrete metric is rather a trivial metric for a space but it often serves as a useful example.

**Definition.** Let (**X**,d) be a metric space and suppose that **A** is a subset of **X**. Let d′ be the map d restricted to $\mathbf{A} \times \mathbf{A}$. Then d′ is called the *induced metric* and (**A**,d′) is called the *induced metric space*.

**Definition.** Let (**X**,d) be a metric space and let $\mathbf{p} \in \mathbf{X}$. The *d-ball of radius r about* ***p***, denoted by $B_r(\mathbf{p},d)$, is defined by

$$B_r(\mathbf{p},d) = \{\mathbf{q} \in \mathbf{X} \mid d(\mathbf{p},\mathbf{q}) < r\}.$$

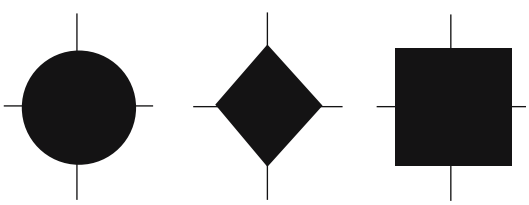
The *d-disk of radius r about* ***p***, denoted by $D_r(\mathbf{p},d)$, is defined by

$$D_r(\mathbf{p},d) = \{\mathbf{q} \in \mathbf{X} \mid d(\mathbf{p},\mathbf{q}) \le r\}.$$

The *d-sphere of radius r about* ***p***, denoted by $S_r(\mathbf{p},d)$, is defined by

$$S_r(\mathbf{p},d) = \{\mathbf{q} \in \mathbf{X} \mid d(\mathbf{p},\mathbf{q}) = r\}.$$

Figure 5.1 shows what the disks of radius 1 around the origin look like in the case of the three metrics defined earlier.



Euclidean metric | taxicab metric | max metric

**Figure 5.1.** Unit disks for different metrics.

**Definition.** Let (**X**,d) be a metric space. A subset **U** of **X** is said to be *d-open* if for every point **p** in **U** there is an $r > 0$ so that $B_r(\mathbf{p},d) \subseteq \mathbf{U}$.

Note that a d-ball is a d-open set. In the case of the discrete metric $\delta$ on a set, every subset is $\delta$-open.

**5.2.3. Theorem.** Let (**X**,d) be a metric space. Then

(1) Both the empty set $\phi$ and the whole space **X** are d-open sets.
(2) An arbitrary union of d-open subsets of **X** is d-open.
(3) Any finite intersection of d-open subsets of **X** is d-open.

**Proof.** The proof is similar to the proof of Proposition 4.2.1.

An **arbitrary** intersection of d-open sets need not be d-open. We already saw an example that shows this for the Euclidean metric d on **R** in Section 4.2.

**Definition.** Let (**X**,d) be a metric space. A subset **C** of **X** is said to be *d-closed* if **X** – **C** is d-open.

A (closed) interval [a,b] in **R** is a closed set with respect to the Euclidean metric. More generally, any d-disk is d-closed, as is the d-sphere.

**5.2.4. Theorem.** Let (**X**,d) be a metric space. Then

(1) Both the empty set $\phi$ and the whole space **X** are d-closed sets.
(2) An arbitrary intersection of d-closed subsets of **X** is d-closed.
(3) Any finite union of d-closed subsets of **X** is d-closed.

**Proof.** See the proof of Proposition 4.2.1.

Again, we already saw in Section 4.2 that an **arbitrary** union of d-closed sets need not be d-closed. Another fact to note is that subsets do not have to be either d-open or d-closed. As was pointed out before, the half-open interval [0,1) in **R** is **neither** d-open **nor** d-closed with respect to the Euclidean metric d.

**Definition.** Let (**X**,d) be a metric space and let $\mathbf{p} \in \mathbf{X}$. A subset **V** of **X** that contains **p** is called a *d-neighborhood of **p*** if there is a d-open set **U** so that $\mathbf{p} \in \mathbf{U} \subseteq \mathbf{V}$.

Note that d-neighborhoods need not be d-open. For example, [–1,1] is a d-neighborhood of 0.

We shall see later in Section 5.3 that the important topological aspect of a space is its collection of open sets. In that sense therefore, although a space can have many different metrics, just because two metrics are different does not mean that the "topology" that they induce on the space is different.

**Definition.** Let d and d′ be two metrics on a space **X**. We shall say that d and d′ are *(topologically) equivalent* metrics if every d-open set of **X** is d′-open and every d′-open set of **X** is d-open.


**Figure 5.2.** Distances between sets.

The Euclidean metric, the taxicab metric, and the max metric on Euclidean space are all equivalent.

**Definition.** Let (**X**,d) be a metric space. Let $\mathbf{p} \in \mathbf{X}$ and $\mathbf{A} \subseteq \mathbf{X}$. Define the *d-distance from **p** to **A***, dist(**p**,**A**), by

$$\mathrm{dist}(\mathbf{p},\mathbf{A}) = \inf\{\mathrm{d}(\mathbf{p},\mathbf{a}) \mid \mathbf{a} \in \mathbf{A}\}.$$

See Figure 5.2(a). One should think of the distance between a point and a set as being the "smallest distance" between the point and points of the set. For example, dist(3,[0,1)) = 2.

**Definition.** Let (**X**,d) be a metric space and let $\mathbf{A},\mathbf{B} \subseteq \mathbf{X}$. Define the *d-distance from **A** to **B***, dist(**A**,**B**), by

$$\mathrm{dist}(\mathbf{A},\mathbf{B}) = \inf\{\mathrm{d}(\mathbf{a},\mathbf{b}) \mid \mathbf{a} \in \mathbf{A}, \mathbf{b} \in \mathbf{B}\}.$$

See Figure 5.2(b). One should think of the distance between two sets as being the "smallest distance" between the points of the sets. For example, dist ([0,1),(3,20]) = 2 and dist ([0,2),(2,3]) = 0.

**Definition.** Let (**X**,d) be a metric space. A subset **A** of **X** is said to be *d-bounded* if there is a constant c > 0 so that d(**p**,**q**) < c for all **p**, **q** ∈ **A**. If the whole space **X** is d-bounded, then the metric space (**X**,d) and the metric d are said to be *bounded*. If **A** is a d-bounded subset of **X**, then the *d-diameter* of **A**, denoted by diam(**A**), is defined by

$$\mathrm{diam}(\mathbf{A}) = \sup\{\mathrm{dist}(\mathbf{p},\mathbf{q}) \mid \mathbf{p},\mathbf{q} \in \mathbf{A}\}.$$

For example, the set **X** = (0,1) is a bounded set in **R**, but the set of integers **Z** is not. The diameter of **X** is 1. The next theorem shows that both a bounded and unbounded metric on a space can induce the same open sets.

**5.2.5. Theorem.** Let (**X**,d) be a metric space. Then **X** admits a bounded metric that is equivalent to d and for which diam(**X**) ≤ 1.

**Proof.** Define

$$\mathrm{d}^*(\mathbf{p},\mathbf{q}) = \min\{1, \mathrm{d}(\mathbf{p},\mathbf{q})\}. \tag{5.5}$$

It is easy to show that d* is a metric on **X** with the desired properties (Exercise 5.2.7). See [Eise74].

**Definition.** Let (**X**,d) and (**Y**,d′) be metric spaces. A one-to-one and onto map $f:\mathbf{X} \to \mathbf{Y}$ is called an *isometry* between these metric spaces if $d'(f(\mathbf{p}),f(\mathbf{q})) = d(\mathbf{p},\mathbf{q})$ for all $\mathbf{p}, \mathbf{q} \in \mathbf{X}$. An arbitrary map $f:\mathbf{X} \to \mathbf{Y}$ is called a *local isometry* if for every **p** in **X** there exist d- and d′-neighborhoods **U** and **V** of **p** and f(**p**), respectively, so that f|**U** is an isometry between **U** and **V**.

Motions in $\mathbf{R}^n$ are all isometries.

**Definition.** Let (**X**,d) and (**Y**,d′) be metric spaces and let $\mathbf{p} \in \mathbf{X}$. A map $f:\mathbf{X} \to \mathbf{Y}$ is said to be *(d,d′)-continuous at **p*** if for every $\varepsilon > 0$ there is a $\delta > 0$ so that $d'(f(\mathbf{p}),f(\mathbf{q})) < \varepsilon$ for all $\mathbf{q} \in \mathbf{X}$ with $d(\mathbf{p},\mathbf{q}) < \delta$. The map f is said to *(d,d′)-continuous* if it is (d,d′)-continuous at every point of **X**.

It is easy to see that in the case of the Euclidean metric this is just the usual definition of continuity of functions $f:\mathbf{R}^n \to \mathbf{R}^m$. See statement (4.1) in Chapter 4.

**Definition.** Let (**X**,d) and (**Y**,d′) be metric spaces. A map $f:\mathbf{X} \to \mathbf{Y}$ is said to be *(d,d′)-uniformly continuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ so that for all $\mathbf{p}, \mathbf{q} \in \mathbf{X}$, $d(\mathbf{p},\mathbf{q}) < \delta$ implies that $d'(f(\mathbf{p}),f(\mathbf{q})) < \varepsilon$.

We already pointed out the big difference between uniform and ordinary continuity in Section 4.2. The $\delta$ in the uniform continuity case does **not** depend on where we are and only on the $\varepsilon$. Having this independence is often important, and so it is always better to have a uniformly continuous map rather than just a plain continuous one.

Next, the usual notion of convergence of sequences in $\mathbf{R}^n$ extends to metric spaces in a natural way.

**Definition.** Let (**X**,d) be a metric space. A sequence of points $\mathbf{p}_n$, n = 1,2, . . . in **X** is said to *converge to a point **p*** in (**X**,d) if for every $\varepsilon > 0$ there is an $m \geq 1$ so that if $n \geq m$ then $d(\mathbf{p}_n,\mathbf{p}) < \varepsilon$. The sequence $\mathbf{p}_n$ is said to *converge* in (**X**,d) if it converges to some point **p** in (**X**,d).

**5.2.6. Theorem.** Let (**X**,d) be a metric space. If a sequence of points $\mathbf{p}_n$, n = 1,2, . . . in **X** converges to two points **p** and **q** in (**X**,d), then **p** = **q**.

**Proof.** Let $\varepsilon > 0$. By definition, there exists an m so that n > m implies that $d(\mathbf{p}_n,\mathbf{p}) < \varepsilon$ and $d(\mathbf{p}_n,\mathbf{q}) < \varepsilon$. But then

$$d(\mathbf{p},\mathbf{q}) \leq d(\mathbf{p},\mathbf{p}_n) + d(\mathbf{p}_n,\mathbf{q}) < \varepsilon + \varepsilon = 2\varepsilon ,$$

which clearly implies that **p** = **q** since $\varepsilon$ was arbitrary.

Theorem 5.2.6 says that if a sequence converges to a point, then it converges to a **unique** point.

**Definition.** If a sequence $\mathbf{p}_n$, n = 1,2, . . . converges to a point $\mathbf{p}$ in a metric space, then this unique point $\mathbf{p}$ is called the *limit point* of the sequence and is denoted by $\lim_{n\to\infty} \mathbf{p}_n$.

**5.2.7. Theorem.** Let (**X**,d) be a metric space. A subset **A** of **X** is d-closed if and only if every limit point in **X** of sequences of points from **A** also belongs to **A**.

**Proof.** If **A** is d-closed, then **X** − **A** is d-open and cannot contain any limit point of **A** since every one of its points has a neighborhood entirely contained in **X** − **A**. The converse is just as easy.

**Definition.** Let (**X**,d) and (**Y**,d′) be metric spaces and consider maps $f_i$, $F:\mathbf{X} \to \mathbf{Y}$. We say that the sequence of maps $f_1, f_2, \ldots$ converges *pointwise* to the map F if, for every $\varepsilon > 0$ and each $\mathbf{p} \in \mathbf{X}$, there is an m so that $d'(f_n(\mathbf{p}),F(\mathbf{p})) < \varepsilon$ for all $n > m$. We say that the sequence $f_1, f_2, \ldots$ converges *uniformly* to F if, for every $\varepsilon > 0$ there is an m so that $d'(f_n(\mathbf{p}),F(\mathbf{p})) < \varepsilon$ for all $n > m$ and all $\mathbf{p} \in \mathbf{X}$. In either case, we call F the *limit function* of the sequence $f_1, f_2, \ldots$.

Notice the important difference between pointwise and uniform convergence. In the former case, the m depends on the ε and the point **p,** and in the latter, it depends **only** on the ε. This is similar to the difference between continuity and uniform continuity. A sequence of functions can converge pointwise but not uniformly (Exercise 5.2.8). The limit function of a sequence of functions, if it exists, is unique because limits of point sequences are unique. One question that arises in the context of sequences of functions is whether the limit function will have a property if all the functions converging to it have this property. The answer to this question is no in general but getting a positive answer in certain cases is precisely why the notion of uniformly convergent is introduced. The next theorem is one example.

**5.2.8. Theorem.** Let (**X**,d) and (**Y**,d′) be metric spaces and consider maps $f_i$, $F:\mathbf{X} \to \mathbf{Y}$. If the maps $f_i$ are continuous and if the sequence of maps $f_1, f_2, \ldots$ converges uniformly to F, then F is continuous.

**Proof.** This is an easy exercise. See [Lips65].

Theorem 5.2.8 is false without the hypothesis of uniform convergence.

**Definition.** Let (**X**,d) be a metric space. A sequence of points $\mathbf{p}_n$, n = 1,2, . . . in **X** is said to be a *Cauchy sequence* in (**X**,d) if for every $\varepsilon > 0$ there is an $m \geq 1$ so that $d(\mathbf{p}_i,\mathbf{p}_j) < \varepsilon$ for all $i,j \geq m$.

**5.2.9. Theorem.** Every convergent sequence in a metric space is a Cauchy sequence.

**Proof.** Let $\mathbf{p}_n$ be a sequence that converges to a point $\mathbf{p}$ in a metric space (**X**,d). Let $\varepsilon > 0$. Choose m so that $n > m$ implies that $d(\mathbf{p}_n,\mathbf{p}) < \varepsilon/2$. It follows that

$$d(\mathbf{p}_i, \mathbf{p}_j) \le d(\mathbf{p}_i, \mathbf{p}) + d(\mathbf{p}, \mathbf{p}_j) < \varepsilon/2 + \varepsilon/2 = \varepsilon ,$$

and the theorem is proved.

**Definition.** A metric space $(\mathbf{X},d)$ is said to be *complete* if every Cauchy sequence in $(\mathbf{X},d)$ converges.

For example, $\mathbf{R}^n$ is complete. The set of rational numbers is **not** complete. The rational numbers have "holes," which is why one defined the real numbers. They fill in those holes.

Finally, taking the product of sets is a common operation. Given a finite collection $(\mathbf{X}_i,d_i)$, $1 \le i \le k$, of metric spaces, we can endow the product space $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$ with many metrics. One natural one that is a generalization of the max metric we defined earlier on $\mathbf{R}^n$ is the following: Let $\mathbf{p} = (\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_k)$, $\mathbf{q} = (\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_k) \in \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$ and define

$$d_\infty(\mathbf{p},\mathbf{q}) = \max_{1\le i\le n}\{d_i(\mathbf{p}_i,\mathbf{q}_i)\} .$$

**Definition.** The function $d_\infty$ is called the *max metric* on $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$.

**5.2.10. Theorem.** Let $(\mathbf{X}_i,d_i)$, $1 \le i \le k$, be metric spaces.

(1) The max metric $d_\infty$ is a metric on the product $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$.
(2) The open sets defined by $d_\infty$ consist of unions of sets of the form $\mathbf{U}_1 \times \mathbf{U}_2 \times \ldots \times \mathbf{U}_k$, where $\mathbf{U}_i$ is $d_i$-open in $\mathbf{X}_i$.

**Proof.** See [Eise74].

In the future, when there is no confusion as to which metric we are talking about we shall drop the "d-" in front of adjectives such as "d-open," "d-closed," etc., and simply say "open," "closed," etc., respectively.

## 5.3 Topological Spaces

As one looks over the topics covered in the last section, one may have noticed the importance of open sets. By in large, the only role the metric played was in defining these sets, but once they were defined the metric was not needed. This leads us to the next level of abstraction where one concentrates on the open sets right from the start. The properties that open sets satisfy in the case of metric spaces (see Theorem 5.2.3) suggest the following:

**Definition.** Let $\mathbf{X}$ be a set. A *topology* on $\mathbf{X}$ is a set T of subsets of $\mathbf{X}$ satisfying

(1) Both $\phi$ and $\mathbf{X}$ belong to T.
(2) The union of **any** collection of sets from T also belongs to T.
(3) The intersection of any **finite** collection of sets from T belongs to T.

The sets in T will be called the *T-open* sets of **X** with respect to the topology T. If T is clear from the context, then we shall simply refer to them as the *open* sets of **X**.

Using induction it is easy to show that condition (3) in the definition of a topology can be replaced by

(3′) The intersection of any **two** sets from T belongs to T.

**Definition.** Let **X** be a set. The *discrete topology* on **X** is the set T of **all** subsets of **X**.

Clearly, the discrete topology on a set **X** is defined by the fact that {**p**} is an open set for each point $\mathbf{p} \in \mathbf{X}$.

**Definition.** A *topological space* is a pair (**X**,T) where **X** is a set and T is a topology on **X**. Again, if T is clear from the context, then one drops the reference to T and simply says the "topological space" **X**.

Intuitively, when one talks about a topological space, what one is saying is that one has specified the collection of subsets that will be called "open." It must be the case though that the empty set and the whole space are open sets, that any union of open sets is open, and that the finite intersection of open sets is open.

**5.3.1. Example.** If (**X**,d) is a metric space, then (**X**,T) is a topological space, where T is set of d-open sets. The topology T is called the *topology on **X** induced by d*. In particular, any subset of $\mathbf{R}^n$ has a topology induced by the standard Euclidean metric.

**5.3.2. Example.** The Euclidean, taxicab, and max metric on $\mathbf{R}^n$ induce the same topology on $\mathbf{R}^n$ or any of its subsets (Exercise 5.3.1). This topology will be called the *standard topology on $\boldsymbol{R}^n$*.

**5.3.3. Example.** The topology on $\mathbf{R}^n$ induced by the discrete metric is the discrete topology and differs from the standard topology. A less trivial example of distinct topologies on a space are the topologies on the space of functions $C^0([0,1])$ induced by the metrics $d_1$ and $d_\infty$ defined by equations (5.2) and (5.3), respectively. For a proof of this fact see [Eise74].

The following is a useful concept.

**Definition.** Let (**X**,T) be a topological space. A *base* for the topology T is a collection of subsets of **X**, so that each subset belongs to T and every set in T is a union of elements of the collection.

For example, the open disks $\mathbf{B}^n(\mathbf{p},r)$ in $\mathbf{R}^n$ are a base for the standard topology on $\mathbf{R}^n$. When dealing with open sets of a topology it often suffices to look at elements of a base. One can also define topologies by means of bases.

**5.3.4. Theorem.** Let $\mathbf{X}$ be a set and $\Omega$ a collection of subsets of $\mathbf{X}$ satisfying:

(1) Each element of $\mathbf{X}$ belongs to some subset of $\Omega$.
(2) If $\mathbf{O}_1, \mathbf{O}_2 \in \Omega$ and $\mathbf{x} \in \mathbf{O}_1 \cap \mathbf{O}_2$, then there exists an $\mathbf{O} \in \Omega$ with $\mathbf{x} \in \mathbf{O} \subseteq \mathbf{O}_1 \cap \mathbf{O}_2$.

Then there is a unique topology T on $\mathbf{X}$ for which $\Omega$ is a base.

**Proof.** See [Eise74].

**Definition.** Let $(\mathbf{X},\mathrm{T})$ be a topological space. A subset $\mathbf{A}$ of $\mathbf{X}$ is called *T-closed* (or simply *closed* if T is clear from the context), if $\mathbf{X} - \mathbf{A}$ is T-open.

**5.3.5. Theorem.** Let $(\mathbf{X},\mathrm{T})$ be a topological space. Then

(1) Both the empty set $\phi$ and the whole space $\mathbf{X}$ are closed sets.
(2) An arbitrary intersection of closed subsets of $\mathbf{X}$ is closed.
(3) Any finite union of closed subsets of $\mathbf{X}$ is closed.

**Proof.** This is easy. See the proof of Theorem 5.2.4.

**Definition.** Let $\mathbf{X}$ be a topological space and let $\mathbf{p} \in \mathbf{X}$. A subset $\mathbf{V}$ of $\mathbf{X}$ that contains $\mathbf{p}$ is called a *neighborhood of* $\boldsymbol{p}$ if there is a open set $\mathbf{U}$ so that $\mathbf{p} \in \mathbf{U} \subseteq \mathbf{V}$. A collection of neighborhoods of $\mathbf{p}$ is called a *local base at* $\boldsymbol{p}$ or a *neighborhood base at* $\boldsymbol{p}$ if every neighborhood of $\mathbf{p}$ contains a member of this collection.

**Definition.** A topological space $(\mathbf{X},\mathrm{T})$ is said to be *metrizable* if $\mathbf{X}$ admits a metric d so that T is the topology on $\mathbf{X}$ induced by d.

Not all topological space are metrizable.

**5.3.6. Example.** Let $\mathbf{X} = \{\mathbf{p},\mathbf{q}\}$ and $\mathrm{T} = \{\phi, \mathbf{X}, \{\mathbf{p}\}, \{\mathbf{p},\mathbf{q}\}\}$. Then $(\mathbf{X},\mathrm{T})$ is a nonmetrizable topological space. If it were metrizable, then $\mathbf{p}$ and $\mathbf{q}$ would have disjoint neighborhoods, which is not the case here.

For less trivial examples of nonmetrizable spaces see [Eise74]. Any space that can be "drawn" (and whose topology is reflected by the picture) will be metrizable because it lives in $\mathbf{R}^n$.

**Definition.** A topological space is said to be a *Hausdorff* space if any two distinct points of $\mathbf{X}$ have disjoint neighborhoods.

The Hausdorff separability condition may seem like an obvious condition that should always hold but it does not follow from our definition of a topology. It is certainly satisfied for metrizable spaces and it is satisfied by all the topological spaces we shall look at, but, as we just saw in Example 5.3.6, not every topological space is a Hausdorff space. Most of the time, however, it is an important technical condition that is assumed for some results to hold. We shall run into it later in quite a few places.

**Definition.** Let **A** be a subset of a topological space **X**. A *boundary point* of **A** is a point **p** in **X** such that every neighborhood of **p** meets both **A** and the complement of **A**, **X** – **A**. The *boundary* of **A** (in **X**), denoted by bdry(**A**), is defined to be the set of boundary points of **A**. A point **a** in **A** is called an *interior point* of **A** (in **X**) if **a** has a neighborhood **U** in **X** that is contained in **A**. The set of interior points of **A** is called the *interior* of **A** and is denoted by int(**A**). The *closure* of **A** (in **X**), denoted by cls(**A**), is defined to be the set of all points **p** in **X** with the property that every neighborhood of **p** meets **A**.

These definitions are analogous to those in Section 4.2. It is important to note that the above definitions of boundary, interior, and closure are all **relative** to the containing space **X**. One can show (Exercise 5.3.2) that

$$\mathrm{int}(\mathbf{A}) = \mathbf{A} - \mathrm{bdry}(\mathbf{A}), \tag{5.6}$$

$$\mathrm{cls}(\mathbf{A}) = \mathrm{int}(\mathbf{A}) \cup \mathrm{bdry}(\mathbf{A}). \tag{5.7}$$

**Definition.** A subset **A** of a topological space **X** is said to be *dense* in **X** if it intersects every nonempty open set of **X**. It is said to be *nowhere dense* in **X** if its closure contains no nonempty open subset of **X**.

For example, the rational numbers are dense in **R**, as are the irrational numbers. Any finite subset of $\mathbf{R}^n$ is nowhere dense in $\mathbf{R}^n$.

**5.3.7. Lemma.** Let (**X**,T) be a topological space and let **Y** be a subset of **X**. Define a set S of subsets of **Y** by

$$\mathrm{S} = \{\mathbf{A} \cap \mathbf{Y} \mid \mathbf{A} \in \mathbf{T}\}. \tag{5.8}$$

Then S is a topology on **Y**.

**Proof.** The proof is straightforward.

**Definition.** Let (**X**,T) be a topological space and let **Y** be a subset of **X**. The topology S on **Y** defined by equation (5.8) is called the *relative topology on **Y** induced by T*. The topological space (**Y**,S) is called a *subspace of (**X**,T)*. If T is clear from the context, one simply talks about the *subspace* **Y** of **X**.

"Induced metrics" and "relative topologies" may sound abstract, but these notions tend to get used automatically in the context of subspaces of $\mathbf{R}^n$, although without explicitly using those terms. For example, one probably had no hesitation to talk about the distance between points **of a circle** or open arcs **in a circle**.

**5.3.8. Lemma.** Let **X** be a topological space and let **Y** be a subspace of **X**. The closed sets of **Y** are just the sets $\mathbf{A} \cap \mathbf{Y}$, where **A** is closed in **X**.

**Proof.** This is an easy consequence of the definitions.

**Definition.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a map from a topological space $\mathbf{X}$ to a topological space $\mathbf{Y}$. Let $\mathbf{p} \in \mathbf{X}$. The map f is said to be *continuous at* $\boldsymbol{p}$ if for every neighborhood $\mathbf{N}$ of $f(\mathbf{p})$ in $\mathbf{Y}$ the set $f^{-1}(\mathbf{N})$ is a neighborhood of $\mathbf{p}$ in $\mathbf{X}$. The map f is said to be *continuous* if it is continuous at every point of $\mathbf{X}$.

**5.3.9. Theorem.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a map from a topological space $\mathbf{X}$ to a topological space $\mathbf{Y}$. The map f is continuous if and only if $f^{-1}(\mathbf{V})$ is open in $\mathbf{X}$ for all open sets $\mathbf{V}$ in $\mathbf{Y}$.

**Proof.** Easy.

We already pointed out in Section 4.2 that in the case of Euclidean spaces the definition of continuity agrees with the standard epsilon-delta definition from calculus. The messy epsilons and deltas obscure the issues that are really at stake. Of course, that definition would actually not be possible here anyway since we do not have a metric. Continuous maps are the natural maps for topological spaces because they involve the only thing that we have in the context of a topology on a set, namely, open sets. Note, however, that we are **not** saying that a continuous map sends open sets to open sets. Such a definition might seem like the obvious one at first glance, but it would not capture what we have in mind.

**Definition.** A map $f: \mathbf{X} \to \mathbf{Y}$ from a topological space $\mathbf{X}$ to a topological space $\mathbf{Y}$ is said to be an *open map* if it maps every open set of $\mathbf{X}$ to an open set of $\mathbf{Y}$. It is said to be a *closed map* if it maps every closed set of $\mathbf{X}$ to an closed set of $\mathbf{Y}$.

The following example shows the difference between a map being continuous, open, and/or closed:

**5.3.10. Example.** Let $\mathbf{X}$ be the reals $\mathbf{R}$ with the standard topology and let $\mathbf{Y}$ be the reals with the discrete topology. Then the identity map from $\mathbf{X}$ to $\mathbf{Y}$ is not continuous but both open and closed. On the other hand, the identity map from $\mathbf{Y}$ to $\mathbf{X}$ is continuous but neither open nor closed.

**5.3.11. Theorem.** If $f: \mathbf{X} \to \mathbf{Y}$ and $g: \mathbf{Y} \to \mathbf{Z}$ are continuous maps between topological spaces, then the composite map $h = f \circ g: \mathbf{X} \to \mathbf{Z}$ is a continuous map.

**Proof.** This follows easily from the definitions.

The next theorem shows that we can piece together continuous maps to get a global continuous map.

**5.3.12. Theorem.** Let $\{\mathbf{A}_i\}_{i \in I}$ be a covering of a topological space $\mathbf{X}$ by subspaces with the property that either all the $\mathbf{A}_i$ are open or all the $\mathbf{A}_i$ are closed and I is finite. If $f: \mathbf{X} \to \mathbf{Y}$ is a map to another topological space, then f is continuous if the restriction maps $f|\mathbf{A}_i$ are continuous for each $i \in I$.

**Proof.** Suppose that the $\mathbf{A}_i$ are open. Let $\mathbf{V}$ be an open set in $\mathbf{Y}$. If $f|\mathbf{A}_i$ is continuous, then

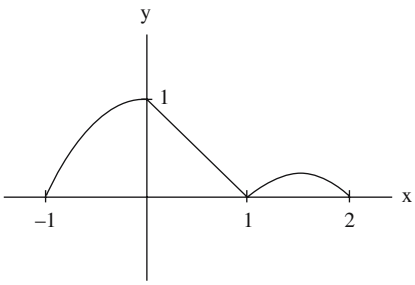



**Figure 5.3.** Piecing together continuous functions.

$$f^{-1}(\mathbf{V}) \cap \mathbf{A}_i = (f \mid \mathbf{A}_i)^{-1}(\mathbf{V})$$

is open in $\mathbf{A}_i$ and $\mathbf{X}$. Therefore,

$$f^{-1}(\mathbf{V}) = f^{-1}(\mathbf{V}) \cap \mathbf{X} = f^{-1}(\mathbf{V}) \cap \bigcup_{i \in I} \mathbf{A}_i = \bigcap_{i \in I} \left(f^{-1}(\mathbf{V}) \cap \mathbf{A}_i\right)$$

is open in $\mathbf{X}$, which proves the continuity of f. The proof of the theorem in the case where the $\mathbf{A}_i$ are closed is also an easy consequence of the definitions. See [Eise74].

**5.3.13. Example.** By Theorem 5.3.12 the map $f:[-1,2] \to \mathbf{R}$ defined by

$$\begin{aligned} f(x) &= -x^2 + 1, && \text{if } x \in [-1,0], \\ &= 1 - x, && \text{if } x \in [0,1], \\ &= -x^2 + 3x - 2, && \text{if } x \in [1,2], \end{aligned}$$

is a continuous map. See Figure 5.3.

**Definition.** Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces. A bijection $f:\mathbf{X} \to \mathbf{Y}$ is called a *homeomorphism* if and only if both f and $f^{-1}$ are continuous maps. Two spaces $\mathbf{X}$ and $\mathbf{Y}$ are said to be *homeomorphic*, and we write $\mathbf{X} \approx \mathbf{Y}$ if there exists a homeomorphism $f:\mathbf{X} \to \mathbf{Y}$. Any property of a space that is preserved by a homeomorphism is called a *topological invariant*.

Homeomorphisms capture the natural notion of equivalence between topological spaces. The identity map of a topological space is a homeomorphism. Homeomorphisms are both open and closed maps. To find a homeomorphism between spaces one first has to find a bijection between them and then needs to show that it preserves open (or closed) sets.

**5.3.14. Example.** Let $f:\mathbf{R} \to \mathbf{R}$ be a continuous map and let

$$\mathbf{X} = \{(x, f(\mathbf{x})) \mid x \in \mathbf{R}\}$$

be the graph of f with the induced topology from $\mathbf{R}^2$. Then the map

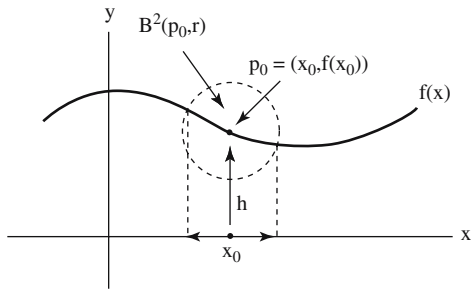



**Figure 5.4.** The domain and graph of a function are homeomorphic.

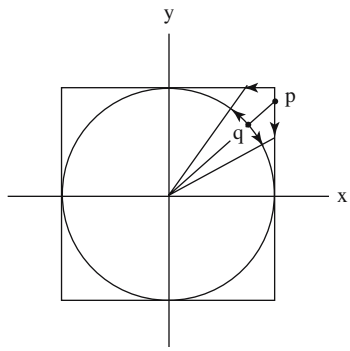



**Figure 5.5.** A circle is homeomorphic to the boundary of a square.

$$h\colon \mathbf{R} \to \mathbf{X} \subset \mathbf{R}^2$$

defined by

$$h(x) = (x, f(x))$$

is a homeomorphism. To see this, let $\mathbf{p}_0 = (x_0, f(x_0))$ and consider an open ball $\mathbf{B}^2(\mathbf{p}_0, r)$ about $\mathbf{p}_0$. See Figure 5.4. Now, $B^2(\mathbf{p}_0, r) \cap \mathbf{X}$ is an open set of $\mathbf{X}$ and it is easy to see that $h^{-1}(\mathbf{B}^2(\mathbf{p}_0, r) \cap \mathbf{X})$ is an open interval in $\mathbf{R}$ containing $x_0$. This and the fact that h is obviously a bijection clearly imply that h is a homeomorphism.

**5.3.15. Example.** The unit circle $\mathbf{S}^1$ is homeomorphic to the boundary $\mathbf{X}$ of the square $[-1,1] \times [-1,1]$. Define a map

$$h\colon \mathbf{X} \to \mathbf{S}^1$$

by

$$h(\mathbf{p}) = \frac{\mathbf{p}}{|\mathbf{p}|}.$$

It is easy to see from Figure 5.5 that h and $h^{-1}$ maps open sets to open sets. Open arcs in $\mathbf{S}^1$ correspond to open "intervals" in $\mathbf{X}$.

*Topological properties* of spaces are those properties that are "preserved" by homeomorphisms. We shall see a number of examples of topological properties in later sections. A *topologist* is someone who tries to find and analyze topological properties of spaces. A topologist does not distinguish between homeomorphic spaces. An ellipse looks the same as a circle. Roughly speaking, two spaces are the same to a topologist if one can be deformed into the other without any ripping or tearing. The deformations correspond to a one-parameter family of homeomorphism. In laymen's terms, topology is "rubber sheet geometry." At the beginning of the next chapter, Chapter 6, we shall have a lot more to say about the kinds of questions that topology tries to answer.

**Definition.** Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces. A map $f:\mathbf{X} \to \mathbf{Y}$ is called an *imbedding* if the map $f:\mathbf{X} \to f(\mathbf{X})$ is a homeomorphism between $\mathbf{X}$ and the subspace $f(\mathbf{X})$ of $\mathbf{Y}$.

Imbeddings are one-to-one maps by definition.

**5.3.16. Example.** If $\mathbf{A}$ is a subspace of $\mathbf{X}$, then the inclusion map of $\mathbf{A}$ into $\mathbf{X}$ is an imbedding. If one gives the set of rational numbers the **discrete** topology, then the inclusion map of this set into the reals $\mathbf{R}$ with the Euclidean topology is **not** an imbedding.

Next, we give some limit-related definitions.

**Definition.** Let $\mathbf{A}$ be a subset of a topological space $\mathbf{X}$. A point $\mathbf{a} \in \mathbf{A}$ is called an *isolated point* of $\mathbf{A}$ if it has a neighborhood that contains no other point of $\mathbf{A}$ except for $\mathbf{a}$. A point $\mathbf{p} \in \mathbf{X}$ is called a *limit* or *accumulation point* of $\mathbf{A}$ if every neighborhood of $\mathbf{p}$ contains a point of $\mathbf{A}$ different from $\mathbf{p}$.

Clearly, every point of a subset $\mathbf{A}$ in a space $\mathbf{X}$ is either an isolated or a limit point. A limit point of $\mathbf{A}$ that does not belong to $\mathbf{A}$ is a point in the boundary of $\mathbf{A}$.

**5.3.17. Theorem.** Let $\mathbf{A}$ be a subset of a metric space $\mathbf{X}$. Every neighborhood of a limit point of $\mathbf{A}$ contains infinitely many points of $\mathbf{A}$.

**Proof.** Easy.

**Definition.** A sequence of points $\mathbf{p}_n$ of a topological space $\mathbf{X}$ is said to *converge* to the point $\mathbf{p}$ in $\mathbf{X}$ if for each neighborhood $\mathbf{U}$ of $\mathbf{p}$ in $\mathbf{X}$ there is some integer m so that $n \geq m$ implies that $\mathbf{p}_n \in \mathbf{U}$.

Many of the properties of convergence of sequences in metric spaces generalize to the context of convergence in topological spaces, but one typically has to add some additional hypotheses (such as the space being Hausdorff) on the type of topology one has. For example, the continuity of a function f defined on a space $\mathbf{X}$ is not in general equivalent to $f(\mathbf{p}_n)$ converging to $f(\mathbf{p})$ for all **sequences** $\mathbf{p}_n$ that converge to $\mathbf{p}$ (*sequential continuity*). One has to allow something more general than sequences. See [Eise74] and the discussion of *nets*.

**Definition.** (The Countability Axioms) A topological space $\mathbf{X}$ is said to be *first countable* if every point of $\mathbf{X}$ has a countable **local** base. It is said to be *second countable* if it has a countable base.

Second countability is clearly stronger than first countability. For continuity to be equivalent to sequential continuity one needs first countability. Second countability is a desirable property that basically means that one can use induction for constructions. It is also related to another concept.

**Definition.** A topological space is said to be *separable* if it has a countable dense subset.

For example, $\mathbf{R}$ is separable because the rationals that are a countable set are dense.

**5.3.18. Theorem.** A metrizable space is separable if and only if it is second countable.

**Proof.** See [Eise74].

We finish this section with a definition of one of the most important types of nice topological spaces, namely, manifolds. Manifolds are really the center of attention of this book. They are basically spaces that look like Euclidean space locally, but because we want to allow for manifolds with boundary the definition is slightly more complicated.

**Definition.** A second countable Hausdorff space $\mathbf{M}$ is called an *n-dimensional topological manifold* if every point $\mathbf{p} \in \mathbf{M}$ has an open neighborhood $\mathbf{V_p}$ that is homeomorphic to an open subset $\mathbf{U_p}$ of the standard halfplane $\mathbf{R}^n_+$. Let $h_\mathbf{p}: \mathbf{U_p} \to \mathbf{V_p}$ be the homeomorphism. The *boundary* of $\mathbf{M}$, $\partial\mathbf{M}$, is defined by

$$\partial\mathbf{M} = \left\{\mathbf{p} \in \mathbf{M} \mid h_\mathbf{p}^{-1}(\mathbf{p}) \in \mathbf{R}^{n-1}\right\}.$$

The points of $\partial\mathbf{M}$ are called *boundary points*. The set $\mathbf{M}-\partial\mathbf{M}$ is called the *interior* of $\mathbf{M}$ and its points are called *interior points*. If n is unimportant one calls $\mathbf{M}$ simply a *topological manifold*. The dimension of $\mathbf{M}$ is usually indicated as a superscript and one talks about the "manifold $\mathbf{M}^n$." A manifold that has no boundary is said to be *closed*.

See Figure 5.6. Euclidean space $\mathbf{R}^n$ is the archetypical example of an n-dimensional manifold without boundary. Other well-known examples are the open balls $\mathbf{B}^n$ and the spheres $\mathbf{S}^n$. The spheres are closed manifolds. The halfplane $\mathbf{R}^n_+$ and the disk $\mathbf{D}^n$ are the archetypical examples of n-dimensional manifolds with boundary. Their boundaries are $\mathbf{R}^{n-1}$ and the (n – 1)-sphere $\mathbf{S}^{n-1}$, respectively. We shall see many more examples of manifolds in coming chapters. A simple-minded way of thinking about a closed two-dimensional manifold is as a space with the property that we can lay a blanket around every point $\mathbf{p}$ of the space so that the points of the blanket match is a one-to-one and onto manner the points of a neighborhood of $\mathbf{p}$.

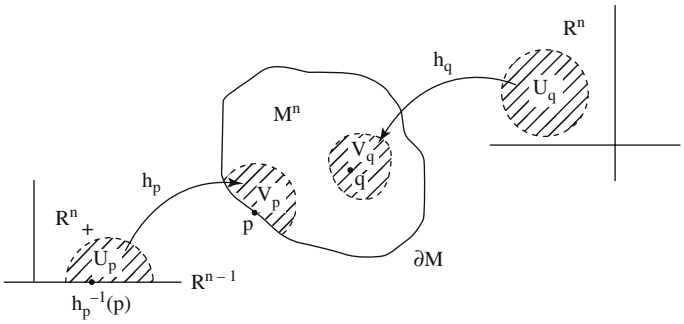



**Figure 5.6.** Neighborhoods of boundary and interior points of a manifold.

We shall show later (Corollary 7.2.3.9) that both the dimension and boundary of a manifold are well defined and do not depend on the neighborhoods $\mathbf{V_p}$ or the homeomorphisms $h_{\mathbf{p}}$. Boundary points are clearly different from interior points. This is easy to see in the one-dimensional case. Consider the interval $\mathbf{I} = [0,1]$, which is a one-dimensional manifold with boundary. Removing a boundary point such as 0 does not disconnect the space, but removing any interior point would. The following facts are easily proved:

(1) Every point of a manifold without boundary has a neighborhood homeomorphic to $\mathbf{R}^n$.
(2) The boundary of an n-dimensional manifold is an (n – 1)-dimensional manifold without boundary.

Because there are other types of manifolds (Chapter 8 will introduce *diffentiable* manifolds and there are also *piecewise linear* or *PL* manifolds) we shall often drop the adjective "topological" and simply refer to a "manifold." The context will always determine the type if it is important.

Some definitions of topological manifolds do not require second countability. The reason for requiring a manifold to be second countable in this book is a practical one. Without it we would lose some properties of manifolds, such as metrizability, and many important results in differential topology described in Chapter 8, such as the Whitney imbedding theorem and the Sard theorem, would no longer hold.

**Definition.** A (*topological*) *surface* is a two-dimensional topological manifold. A (*topological*) *curve* is a one-dimensional topological manifold.

## 5.4 Constructing New Topological Spaces

This section describes some standard construction with which one can define new topological spaces from old ones. The first of these is the important concept of quotient spaces.

**Definition.** Let **X** be a topological space and let ~ be an equivalence relation on **X**. The map

$$\begin{aligned} p\colon\ \mathbf{X} &\to \mathbf{X}/\sim \\ \mathbf{p} &\to [\mathbf{p}] \end{aligned}$$

which sends an element **p** into its equivalence class [**p**] with respect to ~ is called the *quotient map*.

**5.4.1. Lemma.** Let ~ be an equivalence relation on a topological space **X** and let p: **X** → **X**/~ be the quotient map. Define a collection S of subsets of **X**/~ by

$$S = \{\mathbf{B} \subseteq \mathbf{X}/\sim \ \mid\ p^{-1}(\mathbf{B}) \text{ is open in } \mathbf{X}\}. \tag{5.9}$$

Then S is a topology on **X**/~.

**Proof.** The proof is straightforward. See [Eise74].

**Definition.** Let **X** be a topological space and let ~ be an equivalence relation on **X**. The topology S on **X**/~ defined by equation (5.9), is called the *quotient topology* on **X**/~ and the topological space (**X**/~,S) is called a *quotient space* of **X**.

**5.4.2. Lemma.** Let **X** be a topological space and let ~ be an equivalence relation on **X**. The quotient map $p\colon \mathbf{X} \to \mathbf{X}/\sim$ is a continuous map with respect to the quotient topology on **X**/~.

**Proof.** This follows easily from the definition.

**5.4.3. Example.** The Moebius strip can be thought of as the quotient space [0,1] × [0,1]/~, where we use the equivalence relation generated by the relation (0,t) ~ (1,1–t), t ∈ [0,1], between the points of the left and right side of the rectangle. See Figure 5.7.

The next theorem lists some basic properties of quotient spaces.

**5.4.4. Theorem.** Let ~ be an equivalence relation on a topological space **X** and let $p\colon \mathbf{X} \to \mathbf{X}/\sim$ be the quotient map. Let **Y** be a topological space.

(1) A map $g\colon \mathbf{X}/\sim \to \mathbf{Y}$ is continuous if and only if the composite $f = g \circ p\colon \mathbf{X} \to \mathbf{Y}$ is continuous.
(2) If $f\colon \mathbf{X} \to \mathbf{Y}$ is a continuous map that is constant on the equivalence classes of ~, then there is a unique continuous map $f^*\colon \mathbf{X}/\sim \to \mathbf{Y}$ so that $f = f^* \circ p$. The map $f^*$ is called the *induced* (by f) *map* on the quotient space.

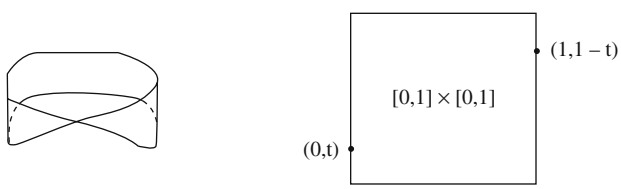



**Figure 5.7.** The Moebius strip.

**Proof.** The proofs are straightforward. See [Eise74]. A good way to remember this theorem is in terms of commutative diagrams. Part (1) says that the map g in the diagram

$$\begin{tikzcd} X \arrow[d, "p"'] \arrow[dr, "f = g \circ p"] & \\ X/\sim \arrow[r, "g"'] & Y \end{tikzcd}$$

is a continuous map if and only it lifts to a continuous map f. Part (2) says that any continuous map f in the diagram

$$\begin{tikzcd} X \arrow[d, "p"'] \arrow[dr, "f"] & \\ X/\sim \arrow[r, "f^*"'] & Y \end{tikzcd}$$

that is constant on equivalence classes induces a unique f*.

It is convenient to introduce some notation for a common special case of a quotient space.

**Definition.** Let **A** be a nonempty subspace of a topological space **X**. Let $\sim_{\mathbf{A}}$ be the equivalence relation

$$\sim_{\mathbf{A}} = \mathbf{A} \times \mathbf{A} \cup \{(\mathbf{p},\mathbf{p}) \mid \mathbf{p} \in \mathbf{X}\}.$$

The quotient space $\mathbf{X}/\sim_{\mathbf{A}}$ will be denoted by **X**/**A** and is usually referred to as the space obtained from **X** by *collapsing **A** to a point*.

The space **X**/**A** is the space we get by identifying all the points of **A** to a single point **A**, the equivalence class of some $\mathbf{a} \in \mathbf{A}$.

**5.4.5. Theorem.** Let **A** be a nonempty subspace of a topological space **X**. If **A** is open or closed, then the quotient map sends $\mathbf{X} - \mathbf{A}$ homeomorphically onto $(\mathbf{X}/\mathbf{A}) - \mathbf{A}$.

**Proof.** Easy.

Let **X** and **Y** be topological spaces. Let **B** be a subspace of **Y** and let $f: \mathbf{B} \to \mathbf{X}$ be a continuous map. We would like to define the space that, intuitively, is obtained from the disjoint union **Z** of **X** and **Y** where we have "attached" or "glued" each point **b** in **B** to f(**b**) in **X**. See Figure 5.8. The identification of **b** with f(**b**) defines an equiv-

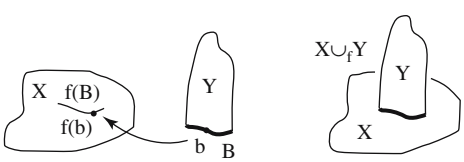



**Figure 5.8.** Attaching one space to another.

alence relation on **Z** and the space we have in mind is just the associated quotient space. To make this precise we need to deal with some technical details. We have to define a disjoint union operation that will handle the case where **X** and **Y** are not disjoint.

**Definition.** The *disjoint union* of two topological spaces **X** and **Y**, denoted by **X** + **Y**, is defined to be the topological space consisting of the set

$$\mathbf{X}+\mathbf{Y}=\mathbf{X}\times\mathbf{0}\cup\mathbf{Y}\times\mathbf{1}$$

and the topology whose open sets are

$$\{\mathbf{U}\times 0\cup\mathbf{V}\times 1 \mid \mathbf{U} \text{ is open in } \mathbf{X} \text{ and } \mathbf{V} \text{ is open in } \mathbf{Y}\}.$$

The spaces **X** and **Y** will always be considered as subspaces of **X** + **Y** under the natural identifications of $\mathbf{x}\in\mathbf{X}$ with $(\mathbf{x},0)$ and $\mathbf{y}\in\mathbf{Y}$ with $(\mathbf{y},1)$.

It is easy to check that the open sets of **X** + **Y** do form a topology, so that we do have a topological space and subspaces **X** and **Y**.

Returning to our map $f:\mathbf{B}\to\mathbf{X}$, we can use it to define an equivalence relation $\sim_f$ on **X** + **Y**.

**Definition.** Let $\mathbf{B}\subseteq\mathbf{Y}$ and let $f:\mathbf{B}\to\mathbf{X}$ be a continuous map. Let $\sim_f$ be the equivalence relation on **X** + **Y** induced by the pairs $(\mathbf{b},f(\mathbf{b}))$, $\mathbf{b}\in\mathbf{B}$. Define

$$\mathbf{X}\cup_f\mathbf{Y}=(\mathbf{X}+\mathbf{Y})/\sim_f.$$

We say that $\mathbf{X}\cup_f\mathbf{Y}$ is obtained from **X** by *attaching **Y** by f* and call the map f the *attaching map*.

**5.4.6. Theorem.** Let $p:\mathbf{X}+\mathbf{Y}\to\mathbf{X}\cup_f\mathbf{Y}$ be the quotient map.

(1) $p(\mathbf{Y}-\mathbf{B})$ is open in $\mathbf{X}\cup_f\mathbf{Y}$ and p maps $\mathbf{Y}-\mathbf{B}$ homeomorphically onto $p(\mathbf{Y}-\mathbf{B})$.
(2) $p(\mathbf{X})$ is closed in $\mathbf{X}\cup_f\mathbf{Y}$ and p maps **X** homeomorphically onto $p(\mathbf{X})$.

**Proof.** See [Eise74].

**5.4.7. Example.** Let **D** be the unit disk $\mathbf{D}^2$, **H** the rectangle $[-1,1]\times[-1,1]$, and **B** the left and right ends $-1\times[-1,1]\cup 1\times[-1,1]$ of **H**. Define $f:\mathbf{B}\to\mathbf{D}$ by

$$f(-1,t)=\left(-\cos\frac{\pi}{6}t,\sin\frac{\pi}{6}t\right)\quad\text{and}\quad f(1,t)=\left(\cos\frac{\pi}{6}t,\sin\frac{\pi}{6}t\right).$$

Then $\mathbf{D}\cup_f\mathbf{H}$ is topologically (homeomorphic to) a disk with a handle. See Figure 5.9.

Sometimes one has a collection of subsets of a set that already have a topology and one wants to extend these topologies to a topology of the whole set. We shall see examples of this in the next chapter.

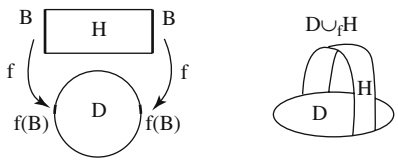



**Figure 5.9.** Attaching a handle to a disk.

**Definition.** Let $\mathbf{X}$ be a set and let $\mathbf{A}_i$ be subsets of $\mathbf{X}$ that already possess a topology (we allow either a finite or infinite set of indices i). Assume

(1) the topologies of $\mathbf{A}_i$ and $\mathbf{A}_j$ agree on $\mathbf{A}_i \cap \mathbf{A}_j$, and either
(2) $\mathbf{A}_i \cap \mathbf{A}_j$ is always open in both $\mathbf{A}_i$ and $\mathbf{A}_j$,

or

(2′) $\mathbf{A}_i \cap \mathbf{A}_j$ is always closed in $\mathbf{A}_i$ and $\mathbf{A}_j$.

The *weak topology* T on $\mathbf{X}$ determined by the topologies of the spaces $\mathbf{A}_i$ is defined by

$$\mathrm{T} = \{\mathbf{U} \subseteq \mathbf{X} \mid \mathbf{U} \cap \mathbf{A}_i \text{ is open in } \mathbf{A}_i \text{ for all i}\}.$$

**5.4.8. Theorem.** Using the notation in the definition of the weak topology, the following holds:

(1) The weak topology is a topology for $\mathbf{X}$.
(2) A subset $\mathbf{A}$ of $\mathbf{X}$ is closed in the weak topology if and only if $\mathbf{A} \cap \mathbf{A}_i$ is closed for all i.
(3) The subsets $\mathbf{A}_i$ will themselves be open subsets of $\mathbf{X}$ in the weak topology if condition (2) in the definition held and closed subsets if (2′) held.

**Proof.** Easy.

Next, one often wants to take the product of topological spaces. Since we want to end up with a topological space, we need to define a product topology. We shall build on what we know for metric spaces.

**5.4.9. Theorem.** If $\mathbf{X}_i$, $1 \le i \le k$, are topological spaces, then the collection of subsets

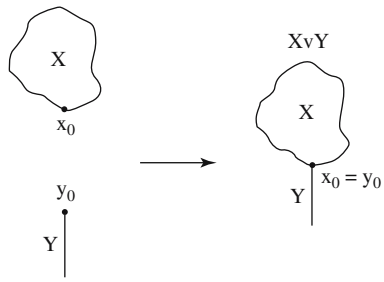
$$\{\mathbf{U}_i \times \mathbf{U}_2 \times \ldots \times \mathbf{U}_k \mid \mathbf{U}_i \text{ is open in } \mathbf{X}_i\}$$

form the base of a unique topology on $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$.

**Proof.** One simply has to show that these subsets satisfy conditions (1) and (2) in Theorem 5.3.4. See [Eise74].

**Definition.** If $\mathbf{X}_i$, $1 \le i \le k$ are topological spaces, then the topology on the product set $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$ described in Theorem 5.4.9 is called the *product topology* on $\mathbf{X}$.

**Figure 5.10.** The wedge of two spaces.




Unless stated otherwise, whenever one takes a product of topological spaces it will always be assumed that the product is given the product topology.

Several other constructions that create new spaces from old are handy.

**Definition.** A *pointed space* is a pair $(\mathbf{X},\mathbf{x}_0)$, where $\mathbf{X}$ is a nonempty topological space and $\mathbf{x}_0 \in \mathbf{X}$. The point $\mathbf{x}_0$ is called the *base point* of the pointed space. The expression "the pointed space $\mathbf{X}$ with base point $\mathbf{x}_0$" will mean the pointed space $(\mathbf{X},\mathbf{x}_0)$.

**Definition.** Let $\mathbf{X}$ and $\mathbf{Y}$ be pointed spaces with base points $\mathbf{x}_0$ and $\mathbf{y}_0$, respectively. Let ~ be the equivalence relation on $\mathbf{X} + \mathbf{Y}$ induced by the pair $(\mathbf{x}_0,\mathbf{y}_0)$. The *one-point union* or *wedge* of $\mathbf{X}$ and $\mathbf{Y}$, denoted by $\mathbf{X} \vee \mathbf{Y}$, is defined to be the pointed space that consists of the quotient space

$$\mathbf{X} \vee \mathbf{Y} = (\mathbf{X} + \mathbf{Y})/\sim$$

and the point to which $\mathbf{x}_0$ and $\mathbf{y}_0$ get identified.

The space $\mathbf{X} \vee \mathbf{Y}$ is just the disjoint union of $\mathbf{X}$ and $\mathbf{Y}$ where we identify $\mathbf{x}_0$ and $\mathbf{y}_0$. See Figure 5.10.

**Definition.** Let $\mathbf{X}$ be a topological space. Define the *cone on X*, denoted by $\mathrm{C}\mathbf{X}$, by

$$\mathrm{C}\mathbf{X} = \mathbf{X} \times [0,1]/\mathbf{X} \times 1.$$

By identifying $\mathbf{X}$ with $\mathbf{X} \times 0$ in $\mathrm{C}\mathbf{X}$, one always considers $\mathbf{X}$ as contained in $\mathrm{C}\mathbf{X}$.

See Figure 5.11. Exercise 5.4.2 gives a more concrete description of $\mathrm{C}\mathbf{X}$.

**Definition.** Let $\mathbf{X}$ be a topological space. Define the *suspension* of $\mathbf{X}$, denoted by $\mathrm{S}\mathbf{X}$, to be the quotient space

$$\mathrm{S}\mathbf{X} = \mathbf{X} \times [-1,1]/\sim,$$

where ~ is the equivalence relation induced by the relations $(\mathbf{x},-1) \sim (\mathbf{x}',-1)$ and $(\mathbf{x},1) \sim (\mathbf{x}',1)$ for all $\mathbf{x}, \mathbf{x}' \in \mathbf{X}$. By identifying $\mathbf{X}$ with $\mathbf{X} \times 0$ in $\mathrm{S}\mathbf{X}$, one always considers $\mathbf{X}$ as contained in $\mathrm{S}\mathbf{X}$.

Again see Figure 5.11. Exercise 5.4.4 gives a more concrete description of $\mathrm{S}\mathbf{X}$.

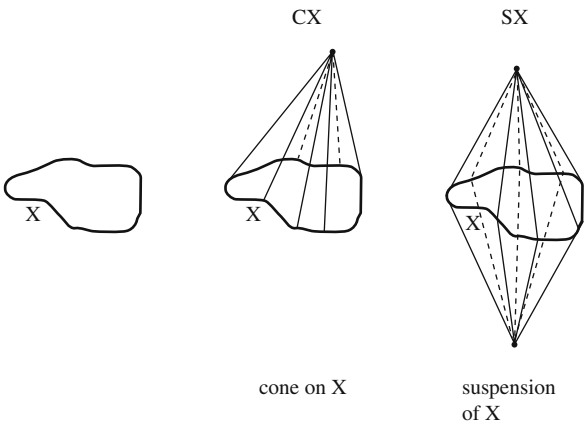



**Figure 5.11.** The cone and suspension of a space.

## 5.5 Compactness

We already pointed out the importance of the compactness property for subsets of $\mathbf{R}^n$ in Chapter 4. The generalization to topological spaces is immediate since its definition was in terms of open sets.

**Definition.** Let $\mathbf{X}$ be a set and let $\mathbf{A} \subseteq \mathbf{X}$. A collection $\Omega$ of subsets of $\mathbf{X}$ is said to be a *cover* of $\mathbf{A}$ if every element of $\mathbf{A}$ belongs to some subset in $\Omega$, that is,

$$\mathbf{A} \subseteq \bigcup_{\mathbf{U} \in \Omega} \mathbf{U}.$$

If $\mathbf{X}$ is a topological space, then we shall call $\Omega$ an *open, closed, . . . cover* of $\mathbf{A}$ if every set in $\Omega$ is open, closed, . . . , in $\mathbf{X}$, respectively. We call $\Omega$ a *finite cover* if it is a finite set. A subset of $\Omega$ that covers $\mathbf{A}$ is called a *subcover* of $\Omega$.

**5.5.1. Example.** The collection

$$\{[n, n+1] \mid n \in \mathbf{Z}\}$$

is a closed cover of $\mathbf{R}$ and

$$\left\{\left(\frac{1}{n}, 1-\frac{1}{n}\right) \mid n = 2, 3, \ldots\right\}$$

is an open cover of (0,1).

**Definition.** A topological space $\mathbf{X}$ is said to be *compact* if every open cover of $\mathbf{X}$ contains a finite subcover of $\mathbf{X}$.

Let $\mathbf{A}$ be a subspace of a topological space $\mathbf{X}$. It is easy to show that $\mathbf{A}$, thought of as a topological space on its own without reference to $\mathbf{X}$, is compact if and only if

every open cover of **A** in **X** contains a finite subcover of **A** (Exercise 5.5.1). Therefore, when it comes to the compactness of a space, we do not have to distinguish between whether we think of a space by itself or as a subspace of another space.

**5.5.2. Theorem.** A closed subset **A** of a compact space **X** is compact.

**Proof.** Let $\{\mathbf{U}_i\}$ be an open cover of **A**. Since **A** is closed, $\mathbf{X} - \mathbf{A}$ is open and so $\{\mathbf{U}_i\} \cup \{\mathbf{X} - \mathbf{A}\}$ is an open cover of **X**. Since **X** is compact, there is a finite subcover and removing the set $\mathbf{X} - \mathbf{A}$, if it is in this subcover, will give us a finite subset of $\{\mathbf{U}_i\}$ that covers **A**.

**5.5.3. Theorem.** A compact subset **A** of a Hausdorff space **X** is closed.

**Proof.** To show that **A** is closed, we need to show that $\mathbf{X} - \mathbf{A}$ is open. The proof proceeds just like the proof for the case $\mathbf{X} = \mathbf{R}^n$ in Theorem 4.2.4. Let $\mathbf{x} \in \mathbf{X} - \mathbf{A}$. Now **X** is a Hausdorff space. Therefore, for every $\mathbf{a} \in \mathbf{A}$ there is an open neighborhood $\mathbf{U_a}$ and $\mathbf{V_a}$ of **a** and **x**, respectively, such that $\mathbf{U_a} \cap \mathbf{V_a} = \phi$. The collection $\{\mathbf{V_a}\}$ is an open cover of **A**. Since **A** is compact, there is a finite subcover $\{\mathbf{V}_{\mathbf{a}_i}\}_{1 \le i \le k}$. It follows that

$$\mathbf{U} = \bigcap_{i=1}^{k} \mathbf{U}_{\mathbf{a}_i}$$

is an open neighborhood of **x** contained in $\mathbf{X} - \mathbf{A}$.

An extremely important theorem with many consequences is the following.

**5.5.4. Theorem.** (The Tychonoff Product Theorem) The product $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$ of nonempty topological spaces $\mathbf{X}_i$ is compact if and only if each $\mathbf{X}_i$ is compact.

**Proof.** See [Eise74].

The Tychonoff product theorem is also true for infinite products.

The next theorem characterizes compact sets in Euclidean space and was already stated and partially proved in Chapter 4 (Theorem 4.2.4). The proof relies on the following:

**5.5.5. Lemma.** (The Heine-Borel Theorem) A closed interval [a,b] in **R** is compact.

**Proof.** Exercise 5.5.2. See [Eise74].

**5.5.6. Theorem.** A subset **A** of $\mathbf{R}^n$ is compact if and only if it is closed and bounded.

**Proof.** We already proved that compact implies closed and bounded in Theorem 4.2.4. We sketch a proof of the converse here.

Let **A** be a closed and bounded subset of $\mathbf{R}^n$. First, consider the case n = 1. Now a closed and bounded subset **A** in **R** is a closed subset of some interval [a,b]. But [a,b] is compact by the Heine-Borel theorem, therefore, Theorem 5.5.2 implies that **A** is

compact. Next, assume that $n > 1$. Again, since **A** is closed and bounded, **A** is a closed subset of a product of intervals $\mathbf{X} = [a_1,b_1] \times [a_2,b_2] \times \ldots \times [a_n,b_n]$. Each interval $[a_i,b_i]$ is compact by the Heine-Borel theorem and so **X** is compact by the Tychonoff product theorem. Theorem 5.5.2 in turn implies that **A** is compact and Theorem 5.5.6 is proved.

The special case of Theorem 5.5.6 where n is 1 is usually referred to as the Heine-Borel-Lebesgue theorem.

Note that a closed and bounded subset of an arbitrary metric space need not be compact. For example, by Theorem 5.2.5 we can give $\mathbf{R}^n$ a bounded metric that induces the standard topology, but $\mathbf{R}^n$ is obviously not compact. It follows that Euclidean space $\mathbf{R}^n$ is special when it comes to Theorem 5.5.6. Although being closed and bounded is equivalent to being compact for subspaces of $\mathbf{R}^n$, the importance of the latter concept is that it clearly shows that we are dealing with an intrinsic property of a space that has nothing to do with any particular imbedding in $\mathbf{R}^n$.

There is a generalization of Theorem 5.5.6 to metric spaces. It asserts that a subset of a complete metric space (**X**,d) is compact if and only if it is d-closed and "d-totally bounded." See [Eise74]. We also have the following:

**5.5.7. Theorem.** A metrizable space **X** is compact if and only if each infinite subset of **X** has a limit point in **X**. A compact metrizable space is complete.

**Proof.** See [Eise74].

From Theorems 5.5.2 and 5.5.3 we get the classical Bolzano-Weierstrass theorem: Every bounded infinite set of real numbers has a limit point in **R**.

**5.5.8. Theorem.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous map between topological spaces. If **X** is compact, then so is f(**X**).

**Proof.** Every open cover of f(**X**) pulls back to an open cover of **X,** which has a finite subcover so that the corresponding open sets in the cover for f(**X**) provide a finite subcover of f(**X**). See the proof of Theorem 4.2.11.

We shall see that Theorem 5.5.8 has many applications because, for example, lots of well-known spaces are quotient spaces of compact spaces. Here are three corollaries.

**5.5.9. Corollary.** Compactness is a topological property, that is, if **X** is homeomorphic to **Y** and if **X** is compact, then so is **Y**.

**Proof.** Clear.

**5.5.10. Corollary.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous map between topological spaces. Assume that **X** is compact and **Y** is Hausdorff. If f is one-to-one and onto, then f is a homeomorphism.

**Proof.** It clearly suffices to show that f is a closed map. Let **A** be a closed subset of **X**. Theorem 5.5.2 implies that **A** is compact and therefore f(**A**) is compact by Theorem 5.5.8. Finally, Theorem 5.5.3 implies that f(**A**) is closed.

Corollary 5.5.10 is useful because it says that when spaces are compact then it is easier to show that maps are homeomorphisms. Normally, one would have to show that their inverse is continuous, but we do not need to prove that here.

**5.5.11. Corollary.** Let $\mathbf{X}$ be a nonempty compact space and assume that $f:\mathbf{X} \to \mathbf{R}$ is a continuous map. Then f attains both its maximum and minimum value on $\mathbf{X}$, that is, there are $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbf{X}$ so that

$$f(\mathbf{x}_1) \le f(\mathbf{x}) \le f(\mathbf{x}_2)$$

for all $\mathbf{x} \in \mathbf{X}$.

**Proof.** By Theorem 5.5.8, $f(\mathbf{X})$ is a compact subset of $\mathbf{R}$ and hence is closed and bounded by Theorem 5.5.6. Therefore, sup $f(\mathbf{X})$ and inf $f(\mathbf{X})$ belong to $f(\mathbf{X})$, proving the existence of $\mathbf{x}_1$ and $\mathbf{x}_2$.

As another example of why compactness is nice, recall Example 4.2.7, which showed that a continuous function need not be uniformly continuous. The next theorem states that if the domain of the function is compact then this does not happen. The proof of the theorem uses the lemma below that we state separately because it has other applications.

**5.5.12. Lemma.** (Lebesgue Covering Lemma) Let $(\mathbf{X},d)$ be a compact metric space. Given any open cover $\Omega$ of $\mathbf{X}$, there is a $\delta > 0$, so that if $\mathbf{p}, \mathbf{q} \in \mathbf{X}$ and $d(\mathbf{p},\mathbf{q}) < \delta$, then there is a set $\mathbf{U}$ in $\Omega$ that contains both $\mathbf{p}$ and $\mathbf{q}$. (The number $\delta$ is called a *Lebesgue number* for $\Omega$.)

**Proof**. See [Eise74].

**5.5.13. Theorem.** Let $f:\mathbf{X} \to \mathbf{Y}$ be a continuous map between metric spaces. If $\mathbf{X}$ is compact, then f is uniformly continuous.

**Proof**. See [Eise74].

We finish this section with the definition for one well-known construction for "compactifying" certain noncompact spaces.

**Definition.** A Hausdorff space $\mathbf{X}$ is said to be *locally compact* if every point has a compact neighborhood.

Clearly, every compact space is locally compact. $\mathbf{R}^n$ is locally compact but not compact.

**5.5.14. Theorem.** Let $\mathbf{X}$ be a noncompact, locally compact topological space with topology T. Let

$$\mathbf{X}_\infty = \mathbf{X} \cup \{\infty\},$$

where we assume that the point $\infty$ does not belong $\mathbf{X}$. Then the set

$$T_\infty = T \cup \{\mathbf{X}_\infty - \mathbf{K} \mid \mathbf{K} \text{ is a compact subset of } \mathbf{X}\}$$

is a topology for $\mathbf{X}_\infty$ that makes $\mathbf{X}_\infty$ into a compact topological space containing $\mathbf{X}$ as a dense subspace. The space $\mathbf{X}_\infty$ is unique in the sense that if $\mathbf{Y}$ is any compact Hausdorff space that has a point $\mathbf{p}$, so that $\mathbf{Y} - \{\mathbf{p}\}$ is homeomorphic to $\mathbf{X}$, then this homeomorphism extends to a homeomorphism from $\mathbf{Y}$ onto $\mathbf{X}_\infty$, which sends $\mathbf{p}$ to $\infty$.

**Proof.** See [Eise74].

**Definition.** The topological space $(\mathbf{X}_\infty, T_\infty)$ is called the *one-point compactification* of $\mathbf{X}$ and the point $\infty$ is called the *point at infinity* in $\mathbf{X}_\infty$.

The obvious example of a one-point compactification is $\mathbf{S}^n$, which is the one-point compactification of $\mathbf{R}^n$.

## 5.6 Connectedness

Connectedness can be defined in a number of ways. Except for the fact that we are dealing with topological spaces, some of the definitions here will be the same as those in Section 4.2, but we shall begin with the pure topological notion.

**Definition.** A topological space $\mathbf{X}$ is said to be *connected* if $\mathbf{X}$ cannot be written in the form $\mathbf{A} \cup \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are two **nonempty** disjoint open subsets of $\mathbf{X}$.

**5.6.1. Theorem.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous map between topological spaces $\mathbf{X}$ and $\mathbf{Y}$. If $\mathbf{X}$ is connected, then so is $f(\mathbf{X})$.

**Proof.** See [Eise74].

**5.6.2. Corollary.** Connectedness is a topological property.

**5.6.3. Theorem.** Consider a collection of nonempty spaces $\mathbf{X}_i$, $1 \le i \le k$. All the $\mathbf{X}_i$ are connected if and only if the product $\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_k$ is connected.

**Proof.** See [Eise74].

**5.6.4. Theorem.** $\mathbf{R}^n$ is connected.

**Proof.** One first proves that $\mathbf{R}$ is connected and then uses Theorem 5.6.3. See [Eise74].

The next theorem uses the connectedness of $\mathbf{R}$ and generalizes the usual intermediate value theorem learned in calculus.

**5.6.5. Theorem.** (The Intermediate Value Theorem) Let $\mathbf{X}$ be a topological space and $f: \mathbf{X} \to \mathbf{R}$ a continuous map. Assume that $f(\mathbf{x}_1) < f(\mathbf{x}_2)$ for some $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{X}$. If c is a real number so that $f(\mathbf{x}_1) < c < f(\mathbf{x}_2)$, then there is an $\mathbf{x} \in \mathbf{X}$ with $f(\mathbf{x}) = c$.

**Proof.** See [Eise74].

**Definition.** A connected subset of a topological space $\mathbf{X}$ that is not properly contained in any connected subset of $\mathbf{X}$ is called a *component* of $\mathbf{X}$.

A more intuitive way to express the notion of component is to say that a component is a maximal connected subset.

A simpler notion of connected is:

**Definition.** Let $\mathbf{X}$ be a topological space. We say that $\mathbf{X}$ is *path-connected* if for any two points $\mathbf{p}, \mathbf{q} \in \mathbf{X}$, there is a continuous map $f: [0,1] \to \mathbf{X}$ with $f(0) = \mathbf{p}$ and $f(1) = \mathbf{q}$. The map f is called a *path from* $\boldsymbol{p}$ *to* $\boldsymbol{q}$. A maximal path-connected subset of a topological space $\mathbf{X}$ is called a *path-component* of $\mathbf{X}$.

**5.6.6. Theorem.** Let $f: \mathbf{X} \to \mathbf{Y}$ be a continuous map from a path-connected space onto a space $\mathbf{Y}$. Then $\mathbf{Y}$ is path-connected.

**Proof.** See [Eise74].

**5.6.7. Theorem.** A path-connected space is connected.

**Proof.** See [Eise74].

Connected does not imply path-connected in general, so that the notion of path-connected is stronger. For "nice" spaces however these concepts are identical.

**5.6.8. Theorem.** A topological manifold is connected if and only if it is path-connected.

**Proof.** See [Eise74].

## 5.7 Homotopy

We have talked about how topology studies properties of spaces invariant under deformations (rubber sheet geometry). This section studies deformations of mappings.

**Definition.** Let $f, g: \mathbf{X} \to \mathbf{Y}$ be continuous maps. A *homotopy between f and g* is a continuous map

$$h: \mathbf{X} \times [0,1] \to \mathbf{Y}$$

such that $h(\mathbf{x},0) = f(\mathbf{x})$ and $h(\mathbf{x},1) = g(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$. In that case, we shall also say that *f is homotopic to g* and write $f \simeq g$.

If we define $f_t : \mathbf{X} \to \mathbf{Y}$ by $f_t(\mathbf{x}) = h(\mathbf{x},t)$, then we can see that the existence of h is equivalent with a one-parameter family of maps connecting f and g and we can think of h as deforming f into g. See Figure 5.12(a).

**5.7.1. Example.** Consider the maps $f, g : \mathbf{D}^2 \to \mathbf{D}^2$ given by $f(\mathbf{p}) = \mathbf{0}$ and $g(\mathbf{p}) = \mathbf{p}$. The map $h : \mathbf{D}^2 \times [0,1] \to \mathbf{D}^2$ defined by $h(\mathbf{p},t) = t\mathbf{p}$ is a homotopy between them. In other words, the identity map of $\mathbf{D}^2$ is homotopic to a constant map.

**5.7.2. Theorem.** The homotopy relation $\simeq$ is an equivalence relation on the continuous maps from one topological space to another.

**Proof.** We must show that the relation is reflexive, symmetric, and transitive.

Reflexivity: If $f : \mathbf{X} \to \mathbf{Y}$ is a continuous map, then $h : \mathbf{X} \times [0,1] \to \mathbf{Y}$ defined by $h(\mathbf{x},t) = f(\mathbf{x})$ is a homotopy between f and f.

Symmetry: Let $f, g : \mathbf{X} \to \mathbf{Y}$ be continuous maps and assume that $h : \mathbf{X} \times [0,1] \to \mathbf{Y}$ is a homotopy between f and g. Define $k : \mathbf{X} \times [0,1] \to \mathbf{Y}$ by $k(x,t) = h(x,1-t)$. Then k is a homotopy between g and f.

Transitivity: Let $f, g, h : \mathbf{X} \to \mathbf{Y}$ be continuous maps and assume that $\alpha, \beta : \mathbf{X} \times [0,1] \to \mathbf{Y}$ are homotopies between f and g and g and h, respectively. Define $\gamma : \mathbf{X} \times [0,1] \to \mathbf{Y}$ by

$$\begin{aligned} \gamma(x,t) &= \alpha(x,2t), & t \in [0,1/2] \\ &= \beta(x,2t-1), & t \in [1/2,1]. \end{aligned}$$

Then $\gamma$ is a homotopy between f and h.

The theorem is proved.

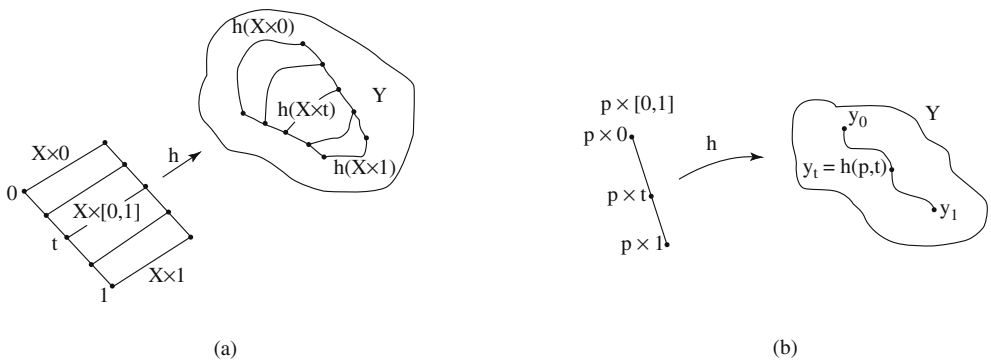



**Figure 5.12.** Homotopies between maps.

**Definition.** If $f:\mathbf{X} \to \mathbf{Y}$ is a continuous map, then the *homotopy class of f*, denoted by [f], is the equivalence class of f with respect to $\simeq$. The set of homotopy classes of maps from $\mathbf{X}$ to $\mathbf{Y}$ will be denoted by $[\mathbf{X},\mathbf{Y}]$.

If $\mathbf{X}$ consists of a single point $\mathbf{p}$, then a homotopy between two maps $f, g:\{\mathbf{p}\} \to \mathbf{Y}$ is just a path in $\mathbf{Y}$ from the point $\mathbf{y}_0 = f(\mathbf{p})$ to the point $\mathbf{y}_1 = g(\mathbf{p})$. See Figure 5.12(b). In particular, it is easy to see that the set of homotopy classes $[\{\mathbf{p}\},\mathbf{Y}]$ is in one-to-one correspondence with the path-components of $\mathbf{Y}$.

**Definition.** A continuous map $f:\mathbf{X} \to \mathbf{Y}$ is called a *homotopy equivalence* if there is a continuous map $g:\mathbf{Y} \to \mathbf{X}$ with $g\circ f \simeq 1_{\mathbf{X}}$ and $f\circ g \simeq 1_{\mathbf{Y}}$. In this case we shall write $\mathbf{X} \simeq \mathbf{Y}$ and say that $\mathbf{X}$ and $\mathbf{Y}$ have the same *homotopy type*.

**5.7.3. Theorem.** Homotopy equivalence is an equivalence relation on topological spaces.

**Proof.** This is straightforward.

Since the general homeomorphism problem is much too difficult except in certain very special cases, a weaker classification is based on homotopy equivalence.

**Definition.** A space is said to be *contractible* if it has the homotopy type of a single point.

**5.7.4. Example.** The unit disk $\mathbf{D}^n$ is contractible. To see this we show that it has the same homotopy type as the point $\mathbf{0}$. Define maps $f:\mathbf{D}^n \to \mathbf{0}$ and $g:\mathbf{0} \to \mathbf{D}^n$ by $f(\mathbf{p}) = \mathbf{0}$ and $g(\mathbf{0}) = \mathbf{0}$. Clearly, $f\circ g = 1_{\mathbf{0}}$. Define $h:\mathbf{D}^n \times [0,1] \to \mathbf{D}^n$ by $h(\mathbf{p},t) = t\mathbf{p}$. Then h is a homotopy between $g\circ f$ and the identity map on $\mathbf{D}^n$, and we are done. Another way to state the result is to say that both f and g are homotopy equivalences.

**Definition.** A subspace $\mathbf{A}$ of a space $\mathbf{X}$ is called a *retract* of $\mathbf{X}$ if there exists a continuous map $r:\mathbf{X} \to \mathbf{A}$ with $r(\mathbf{a}) = \mathbf{a}$ for all $\mathbf{a}$ in $\mathbf{A}$. The map r is called a *retraction of* $\boldsymbol{X}$ *onto* $\boldsymbol{A}$.

If $\mathbf{x}_0$ is any point in a space $\mathbf{X}$, the constant map $r(\mathbf{x}) = \mathbf{x}_0$ shows that any point of a space is a retract of the space. A less trivial example is

**5.7.5. Example.** The unit circle in the plane is a retract of the cylinder

$$\mathbf{X} = \{(x,y,z) \mid X^2 + Y^2 = 1 \text{ and } z \in [0,1]\} \tag{5.10}$$

because we have the retraction $r(x,y,z) = (x,y,0)$.

**Definition.** Let $\mathbf{A}$ be a subspace of a space $\mathbf{X}$. A *deformation retraction of* $\boldsymbol{X}$ *onto* $\boldsymbol{A}$ is a continuous map $h:\mathbf{X} \times \mathbf{I} \to \mathbf{X}$ satisfying

$$h(\mathbf{x},0) = \mathbf{x}, \quad h(\mathbf{x},1) \in \mathbf{A}, \quad \text{all} \quad \mathbf{x} \in \mathbf{X},$$

and

$$h(\mathbf{a},1) = \mathbf{a}, \quad \text{all} \quad \mathbf{a} \in \mathbf{A}.$$

In this case **A** is called a *deformation retract* of **X**.

The argument in Example 5.7.4 also shows that the map $f:\mathbf{D}^n \to \mathbf{0}$ defined by $f(\mathbf{p}) = \mathbf{0}$ is a deformation retraction of $\mathbf{D}^n$ onto $\mathbf{0}$.

**5.7.6. Example.** The unit circle is a deformation retract of the cylinder **X** defined by equation (5.10). To see this simply define $h:\mathbf{X} \times \mathbf{I} \to \mathbf{X}$ by $h((x,y,z),t) = (x,y,(1-t)z)$.

**5.7.7. Theorem.** Let **A** be a subspace of a space **X**. If **A** is a deformation retract of **X**, then the inclusion map $\iota:\mathbf{A} \to \mathbf{X}$ is a homotopy equivalence. In particular, **A** and **X** have the same homotopy type.

**Proof.** Let $h:\mathbf{X}\times\mathbf{I}\to\mathbf{X}$ be a deformation retraction of **X** onto **A**. Define $f:\mathbf{X}\to\mathbf{A}$ by $f(\mathbf{x}) = h(\mathbf{x},1)$. Since $f\circ\iota = 1_{\mathbf{A}}$ and h is homotopy between $\iota\circ f$ and $1_{\mathbf{X}}$, we are done.

Intuitively speaking, a subset **A** is a deformation retract of a space **X** if we can shrink **X** down to **A** without "cutting" anything. We shall see later (Corollary 7.2.3.3 and Theorem 7.2.3.4) that a circle does not have the same homotopy type as a point. Therefore, no point of the circle is a deformation retract of the circle. The only way to "shrink" the circle to a point would be to cut it first.

Often it is convenient to talk about "pointed" homotopies, or more generally "relative homotopies."

**Definition.** The notation $f:(\mathbf{X},\mathbf{A}) \to (\mathbf{Y},\mathbf{B})$ will mean that f is a map from **X** to **Y** and $f(\mathbf{A}) \subseteq \mathbf{B}$.

**Definition.** Let $f, g:(\mathbf{X},\mathbf{A}) \to (\mathbf{Y},\mathbf{B})$ be continuous maps. A *homotopy between f and g relative **A*** is a continuous map

$$h:\mathbf{X}\times[0,1]\to\mathbf{Y}$$

such that $h(\mathbf{x},0) = f(\mathbf{x})$, $h(\mathbf{x},1) = g(\mathbf{x})$, and $h(\mathbf{a},t) \in \mathbf{B}$ for all $\mathbf{x} \in \mathbf{X}$, $\mathbf{a} \in \mathbf{A}$, and $t \in [0,1]$. In that case, we shall also say that *f is homotopic to g relative **A*** and write $f \simeq_{\mathbf{A}} g$.

**5.7.8. Theorem.** The homotopy relation $\simeq_{\mathbf{A}}$ is an equivalence relation on the set of continuous maps $f:(\mathbf{X},\mathbf{A}) \to (\mathbf{Y},\mathbf{B})$.

**Proof.** The proof is similar to the proof of Theorem 5.7.2. We just have to be careful that the homotopies keep sending **A** to **B**.

**Definition.** The set of homotopy classes of maps $f:(\mathbf{X},\mathbf{A}) \to (\mathbf{Y},\mathbf{B})$ with respect to the equivalence relation $\simeq_{\mathbf{A}}$ will be denoted by $[(\mathbf{X},\mathbf{A}),(\mathbf{Y},\mathbf{B})]$.

A natural question to ask at this point is how many homotopy classes of maps there are between spaces in general and what this number measures. We are not ready

to answer such a question yet and will have to wait until Chapter 7, but the reader may appreciate one glimpse into the future. Given a map

$$f:\mathbf{S}^1 \to \mathbf{S}^1,$$

the *degree of f* is, intuitively, the number of times that f winds the circle around itself. (In Section 7.5.1 we shall give another definition of the degree of f.) Define

$$f_n:\mathbf{S}^1 \to \mathbf{S}^1$$

by $f_n(\cos\theta,\sin\theta) = (\cos n\theta,\sin n\theta)$. Then $f_n$ has degree n. It turns out that all maps of the circle to itself are homotopic to one of these maps and two maps are homotopic only if they have the same degree, so that there is a bijection between the homotopy classes of maps of the circle to itself and the integers.

## 5.8 Constructing Continuous Functions

There are many situations where one wants to define continuous functions on a topological space satisfying certain properties. This brief section describes two very fundamental theorems that deal with the existence of certain functions, which in turn can be used to construct many other functions. We shall give one application having to do with the existence of partitions of unity at the end of the section.

For a topological space to have the continuous functions we want it needs to satisfy a special property. It is worth isolating this property and giving it a name.

**Definition.** A topological space **X** is said to be *normal* if, given two disjoint closed sets **A** and **B** in **X**, there exist disjoint open sets containing **A** and **B**, respectively.

The condition that a space be normal is somewhat technical, like being Hausdorff, but fortunately the spaces of interest to us satisfy this property.

**5.8.1. Theorem.**

(1) Any metrizable space is normal.
(2) Any compact Hausdorff space is normal.

**Proof.** See [Eise74].

**5.8.2. Theorem.** (The Urysohn Lemma) Let **X** be a normal space and assume that **A** and **B** are two closed subsets of **X**. Then there exists a continuous function $f:\mathbf{X} \to [0,1]$ such that f takes the value 0 on **A** and 1 on **B**.

**Proof.** See [Jäni84].

**5.8.3. Theorem.** (Tietze Extension Lemma) Let **X** be a normal space. Then any continuous function $f:\mathbf{A} \to [a,b]$ or $f:\mathbf{A} \to \mathbf{R}$ defined on a closed subset **A** of **X** can be extended to a continuous function $F:\mathbf{X} \to [a,b]$ or $F:\mathbf{X} \to \mathbf{R}$, respectively.

**Proof.** See [Jäni84].

The next concept enables one to localize problems and will be used in later chapters.

**Definition.** A *partition of unity* on a topological space **X** is a collection $\Phi$ of continuous real-valued functions satisfying the following:

(1) For all $\varphi \in \Phi$ and $\mathbf{x} \in \mathbf{X}$, $0 \le \varphi(\mathbf{x}) \le 1$.
(2) Every point in **X** has a neighborhood on which all but a finite number of functions in $\Phi$ vanish.
(3) For every **x** in **X**

$$\sum_{\varphi \in \Phi} \varphi(\mathbf{x}) = 1.$$

(Note that by condition (2) this is a finite sum for each **x**.)

If $\xi$ is a cover of **X**, we say that the partition of unity $\Phi$ is *subordinate* to $\xi$ if each function in $\Phi$ vanishes outside some set in $\xi$.

**5.8.4. Example.** Define functions

$$b, b_n : \mathbf{R} \to \mathbf{R}$$

by

$$\begin{aligned} b(x) &= x+1, \quad \text{for} \quad x \in [-1,0], \\ &= -x+1, \quad \text{for} \quad x \in [0,1], \\ &= 0, \quad \text{elsewhere}, \end{aligned}$$

and

$$b_n(x) = b(x - n).$$

It is easy to check that the collection of functions $b_n(x)$ is a partition of unity on **R**. See Figure 5.13. This partition of unity is subordinate to the open cover

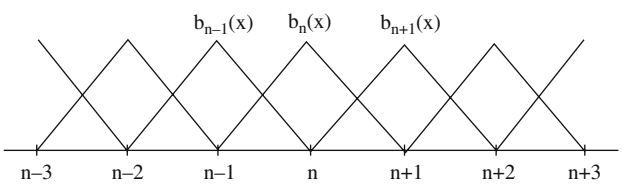



**Figure 5.13.** A partition of unity for **R**.

$$\{(n-1, n+1) \mid n \in \mathbf{Z}\}$$

of **R**.

It would be good to know what condition on a topological space guarantees the existence of partitions of unity.

**Definition.** A cover of a space is *locally finite* if every point in the space has a neighborhood that meets only finitely many elements of the cover. A Hausdorff space is said to be *paracompact* if every open cover admits a locally finite subcover.

**5.8.5. Lemma.** Every paracompact space is normal.

**Proof.** See [Jäni84].

**5.8.6. Theorem.** A Hausdorff space is paracompact if and only if every open cover admits a partition of unity subordinate to it.

**Proof.** The only hard part is showing that paracompact implies the existence of the stated partitions of unity. Because of Lemma 5.8.5 one can use Urysohn's lemma to construct the desired partition of unity. See [Jäni84].

The next obvious question is: which spaces are paracompact?

**5.8.7. Theorem.** The following types of topological spaces are paracompact:

(1) Compact Hausdorff spaces
(2) Topological manifolds
(3) Metrizable spaces

**Proof.** Part (1) is trivial. Part (2) is also not hard. For (3) see [Schu68].

## 5.9 The Topology of $P^n$

Projective space $\mathbf{P}^n$ is not only one of the really important spaces in mathematics but it also serves as an excellent example of a nontrivial topological space. This section looks at its purely topological properties. We shall return to it later in Chapter 8 to look at its manifold properties and again in Chapter 10 where its algebraic properties come to the fore.

Recall the (set theoretic) definition of $\mathbf{P}^n$ given in Section 3.4, namely,

$$\mathbf{P}^n = \left(\mathbf{R}^{n+1} - \mathbf{0}\right)\big/\sim, \tag{5.11}$$

where ~ is the equivalence relation on $\mathbf{R}^{n+1} - \mathbf{0}$ defined by $\mathbf{p} \sim c\mathbf{p}$, for $c \neq 0$. In Chapter 3 we did not say anything about its topology, but actually, when we talk about $\mathbf{P}^n$ as a topological space, we always assume that it has been given the quotient space topology that is defined by equation (5.11). What does this space really "look" like top-

ologically? There are quite a few different definitions that all lead to the same space (up to homeomorphism). Each gives a little different insight into its structure.

**A second definition of $\mathbf{P}^n$:** $\mathbf{P}^n$ is the set of lines through the origin in $\mathbf{R}^{n+1}$.

**Justification:** Except for the fact that the origin is missing, the equivalence class $[x_1, x_2, \ldots, x_{n+1}]$ is just such a line through the origin, so that there is a natural one-to-one correspondence of points. (The topologies are assumed to match under this correspondence.)

**A third definition of $\mathbf{P}^n$:** $\mathbf{P}^n$ is the unit sphere $\mathbf{S}^n$ with antipodal points identified, that is,

$$\mathbf{P}^n = \mathbf{S}^n / \sim,$$

where $\mathbf{p} \sim -\mathbf{p}$.

**Justification:** The relation $\sim$ which relates points of $\mathbf{S}^n$ to their antipodal points is an equivalence relation, and the map

$$\left(\mathbf{R}^{n+1} - \mathbf{0}\right)\big/\sim \to \mathbf{S}^n / \sim,$$

$$[x_1, x_2, \ldots, x_{n+1}] \to \left[\frac{x_1}{D}, \frac{x_2}{D}, \ldots, \frac{x_{n+1}}{D}\right]$$

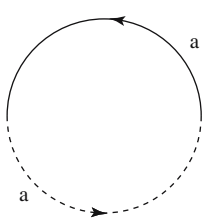
where $D = |(x_1, x_2, \ldots, x_{n+1})|$, is clearly a homeomorphism.

**A fourth definition of $\mathbf{P}^n$:** $\mathbf{P}^n$ is the unit disk $\mathbf{D}^n$ in the plane with antipodal points on its boundary identified, that is,

$$\mathbf{P}^n = \mathbf{D}^n \big/ \sim,$$

where $\sim$ is induced from the relations $\mathbf{p} \sim -\mathbf{p}$ for $\mathbf{p} \in \mathbf{S}^{n-1}$.

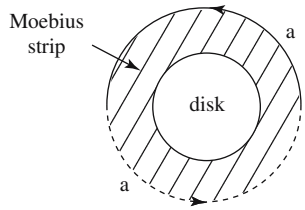
**Justification:** See Figure 5.14 where the labels and arrows are trying to indicate the identifications. The boundary of the upper hemisphere $\mathbf{S}^n_+$ is just $\mathbf{S}^{n-1}$. It is easy to see




disk with antipodal boundary points identified

**Figure 5.14.** Visualizing the projective plane.

**Figure 5.15.** Another way to visualize the projective plane.




Disk glued to Moebius strip along their boundary

that $\mathbf{P}^n$ can be thought of as $\mathbf{S}^n_+/\sim$, where ~ is a restricted version of the equivalence relation in the third definition above, namely, $\mathbf{p} \sim -\mathbf{p}$ for $\mathbf{p} \in \mathbf{S}^{n-1}$. Our characterization now follows from the observation that the only identifications that are taking place are on the boundary of $\mathbf{S}^n_+$ and that the interior of $\mathbf{S}^n_+$ projects in a one-to-one fashion onto the interior of the disk $\mathbf{D}^n$.

In the case of the projective plane, there is another well-known identification.

**A fifth definition of $\mathbf{P}^2$**: The projective plane is the union of the Moebius strip and a disk where we identify their boundaries, which are just circles.

**Justification:** Consider the shaded region in Figure 5.15, which is a "collar" of the boundary of the unit disk. A little thought should convince the reader that under the identification described in the fourth definition, this shaded region is just the Moebius strip.

We now have four different ways of looking at the topological space $\mathbf{P}^n$ (five, in the case of the projective plane). In each case we used a quotient topology of Euclidean space. Alternatively, one can define this topology by defining a metric on $\mathbf{P}^n$.

**Definition.** Let $\mathbf{p} = [\mathbf{x}]$ and $\mathbf{q} = [\mathbf{y}]$ be points of $\mathbf{P}^n$, where $\mathbf{x}, \mathbf{y} \in \mathbf{R}^{n+1}$. Define the *distance between* $\boldsymbol{p}$ *and* $\boldsymbol{q}$, denoted dist $(\mathbf{p},\mathbf{q})$, by

$$\operatorname{dist}(\mathbf{p},\mathbf{q}) = \mathrm{d}, \quad \text{where } \cos \mathrm{d} = \frac{|\mathbf{x} \bullet \mathbf{y}|}{|\mathbf{x}||\mathbf{y}|} \quad \text{and} \quad 0 \le \mathrm{d} \le \pi/2.$$

It is easy to see that dist$(\mathbf{p},\mathbf{q})$ is well defined and does not depend on the representatives $\mathbf{x}$ and $\mathbf{y}$ that are chosen for $\mathbf{p}$ and $\mathbf{q}$, respectively. It is just the angle between the two "lines" $\mathbf{p}$ and $\mathbf{q}$. The function dist$(\mathbf{p},\mathbf{q})$ is in fact a metric on $\mathbf{P}^n$ and makes $\mathbf{P}^n$ into a metric space.

**5.9.1. Example.** Let $\mathbf{L}$ be the line in $\mathbf{R}^2$ defined by $-x + 2y + 1 = 0$. To find the points of $\mathbf{P}^2$ that are "near" the ideal point $\mathbf{L}^\infty$ associated to the family of lines parallel to $\mathbf{L}$.

**Solution.** Let $\mathbf{p} = [X,Y,Z]$ be any point of $\mathbf{P}^2$. Since $\mathbf{L}^\infty = [-2,-1,0]$, a simple computation using the definition of distance shows that the distance d between the two points satisfies

$$\cos d = \frac{|-2X - Y|}{\sqrt{5}\,|(X,Y,Z)|}.$$

If **p** corresponds to a real point (x,y), that is, **p** = [x,y,1], the only way that d will go to zero is if x and y both get arbitrarily large and (x,y) is near the line y – 2x = 0. This follows from the Cauchy-Schwarz inequality. Similarly, the only ideal points close to $\mathbf{L}^{\infty}$ are points **p** = [X,Y,0] with X close to –2 and Y close to –1.

**5.9.2. Theorem.** N-dimensional projective space $\mathbf{P}^n$ is a compact, connected, metrizable topological manifold.

**Proof.** The compactness and connectedness follows from Lemma 5.3.18 and Theorems 5.4.6 and 5.5.2 using the third and fourth definition of $\mathbf{P}^n$. We postpone showing that $\mathbf{P}^n$ is a manifold to Section 8.13, where we will in fact show that it is a differentiable manifold.

Finally, we note that any hyperplane in $\mathbf{P}^n$ is homeomorphic to $\mathbf{P}^{n-1}$. In particular, the subspace of ideal points is homeomorphic to $\mathbf{P}^{n-1}$.

## 5.10 Exercises

### Section 5.2

5.2.1. Prove that every metric on a finite set is the discrete metric.

5.2.2. Prove that if a vector space **V** has an inner product <,>, then the function

$$d(\mathbf{u},\mathbf{v}) = |\mathbf{uv}| = \sqrt{<\mathbf{v}-\mathbf{u},\mathbf{v}-\mathbf{u}>}$$

defines a metric on **V**.

5.2.3. Prove that equation (5.1) defines an inner product on $C^0([0,1])$.

5.2.4. Prove that the function $d_1$ defined by equation (5.3) defines a metric on $C^0([0,1])$.

5.2.5. Prove that the function $d_\infty$ defined by equation (5.4) defines a metric on $C^0([0,1])$.

5.2.6. Show that the metrics $d_1$ and $d_\infty$ on $C^0([0,1])$ defined by equations (5.3) and (5.4), respectively, are not equivalent metrics.

5.2.7. Let (**X**,d) be a metric space. Prove that the function d* defined by equation (5.5) is a bounded metric on **X**.

5.2.8. Consider the sequence of functions $f_n(x) = x^n$ on [0,1]. Show that this sequence of functions converges in a pointwise fashion but not uniformly. Note also that although each function is continuous, the limit function g(x) is not. Describe g(x).

5.2.9. Show that the rational numbers are not complete by giving an example of a Cauchy sequence of rational numbers that does not converge to a rational number.

5.2.10. Sometimes one does not quite have a metric on a space.

**Definition.** A *pseudometric* on a set $\mathbf{X}$ is a function

$$d\colon \mathbf{X} \times \mathbf{X} \to \mathbf{R}$$

satisfying conditions (1), (3), and (4) in the definition of a metric and the following weakened form of condition (2):

(2′) $d(\mathbf{p},\mathbf{p}) = 0$ for all $\mathbf{p} \in \mathbf{X}$.

(a) Show that the relation ~ on $\mathbf{X}$ defined by

$$\mathbf{p} \sim \mathbf{q} \text{ if and only if } d(\mathbf{p},\mathbf{q}) = 0$$

is an equivalence relation on $\mathbf{X}$.

(b) Let $\mathbf{X}^*$ denote the set of equivalence classes of $\mathbf{X}$ with respect to the relation ~ in (a). Define

$$d^*\colon \mathbf{X}^* \times \mathbf{X}^* \to \mathbf{R}$$

by

$$d^*([\mathbf{p}],[\mathbf{q}]) = d(\mathbf{p},\mathbf{q}).$$

Show that $d^*$ is a well-defined metric on $\mathbf{X}^*$.

### Section 5.3

5.3.1. Prove that the Euclidean, taxicab, and max metric on $\mathbf{R}^n$ induce the same topology.

5.3.2. Prove equations (5.6) and (5.7).

5.3.3. Define a homeomorphism between the open interval (0,1) and $\mathbf{R}$.

### Section 5.4

5.4.1. Prove that if the maps $f_i\colon \mathbf{X}_i \to \mathbf{Y}_i$ are continuous, then so is the map

$$f_1 \times f_2 \times \ldots \times f_n\colon X_1 \times X_2 \times \ldots \times X_n \to Y_1 \times Y_2 \times \ldots \times Y_n.$$

5.4.2. We can give a more concrete description of the cone on a subspace of $\mathbf{R}^n$. Let $\mathbf{X} \subseteq \mathbf{R}^n$. Choose $\mathbf{v} \in \mathbf{R}^{n+1} - \mathbf{R}^n$. Prove that the cone $C\mathbf{X}$ on $\mathbf{X}$ is homeomorphic to the space

$$\{t\mathbf{x} + (1-t)\mathbf{v} \mid \mathbf{x} \in \mathbf{X} \text{ and } t \in [0,1]\}.$$

5.4.3. Prove that $\mathbf{D}^n \approx C\mathbf{S}^{n-1}$.

5.4.4. We can give a more concrete description of the suspension of a subspace of $\mathbf{R}^n$. Let $\mathbf{X} \subseteq \mathbf{R}^n$. Choose $\mathbf{v} \in \mathbf{R}^{n+1} - \mathbf{R}^n$ and $\mathbf{w} \in \mathbf{R}^{n+2} - \mathbf{R}^{n+1}$. Prove that the suspension $S\mathbf{X}$ of $\mathbf{X}$ is homeomorphic to the space

$$\{t\mathbf{x} + (1-t)\mathbf{u} \mid \mathbf{x} \in \mathbf{X},\ t \in [0,1], \text{ and } \mathbf{u} = \mathbf{v} \text{ or } \mathbf{w}\}.$$

5.4.5. Prove that $\mathbf{S}^n \approx S\mathbf{S}^{n-1}$.

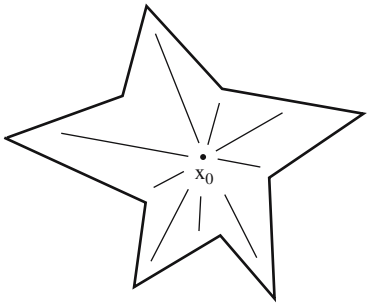



**Figure 5.16.** A star-shaped region.

**Section 5.5**

5.5.1. Let **A** be a subspace of a topological space **X**. Prove that **A** is compact if and only if every cover of **A** by open subsets of **X** has a finite subcover.

5.5.2. Prove that a closed interval [a,b] in **R** is compact.

**Hint:** Consider

$$\mathbf{A} = \{x \mid [0,x] \text{ can be covered by a finite number of sets from the open cover}\}.$$

Prove

(1) $0 \in \mathbf{A}$
(2) $c = \sup \mathbf{A} \in \mathbf{A}$ (Find a contradiction to the assumption that $0 < c \notin \mathbf{A}$ by finding a $\delta > 0$ so that $[0,c-\delta]$ is compact and $[\delta,c]$ is contained in an open set from the cover.)
(3) $c = 1$

**Section 5.6**

5.6.1. Prove that a space consisting of two points is not connected.

5.6.2. Prove that every convex subset of $\mathbf{R}^n$ is connected.

5.6.3. Use a connectivity argument to justify the fact that a "figure eight" (the wedge $\mathbf{S}^1 \vee \mathbf{S}^1$) is not homeomorphic to a circle.

**Section 5.7**

5.7.1. Show that a space **X** is contractible if and only if the identity map for **X** is homotopic to a constant map $g:\mathbf{X} \to \mathbf{X}$.

5.7.2. Let $\mathbf{X} \subseteq \mathbf{R}^n$. Define **X** to be *star-shaped* if there is some point $\mathbf{x}_0 \in \mathbf{X}$, such that for every $\mathbf{x} \in \mathbf{X}$, the segment $[\mathbf{x}_0,\mathbf{x}]$ is contained in **X**. See Figure 5.16. Prove that every star-shaped region is contractible.

5.7.3. Prove that every cone is contractible. This extends the result from Exercise 5.7.1.

5.7.4. Prove that a retract of a contractible space is contractible.

CHAPTER 6

# Combinatorial Topology

## 6.1 Introduction

Topology is a relatively new field in mathematics. In the last chapter we considered some basic concepts from what is called point set topology. This chapter is the first of three that will introduce us to a broader view of the subject and we begin with a brief history. Before the mid-19$^{th}$ century the field as a whole consisted mainly of a collection of isolated facts. The word "topology" itself appeared first in 1847 in the book *Vorstudien zur Topologie*, by J.B. Listing, a student of Gauss. Before that, one used the term "geometria situs" for the study of certain qualitative properties of geometric figures that would be considered topological today. The term was introduced by Leibniz, although he did not contribute much to the subject himself. In the 1800s and early 1900s topology was usually referred to as "analysis situs."

Probably the earliest significant topological observation concerned a relationship between the number of faces, edges, and vertices of a simple polyhedron, which was already known to Descartes around 1620. By a *simple polyhedron* we mean a convex three-dimensional linear polyhedron, that is, a convex solid figure without holes that is bounded by planar faces. The five standard well-known regular simple polyhedra are shown in Figure 6.1. A *regular polyhedron* is a polyhedron with the property that every face has the same number of edges and every vertex has the same number of edges emanating from it. For now we are only interested in the boundary of a simple polyhedron. It is easy to show that a simple polyhedron is homeomorphic to $\mathbf{D}^3$, so that its boundary is homeomorphic to the 2-sphere $\mathbf{S}^2$.

Given a simple polyhedron, let $n_v$, $n_e$, and $n_f$ denote the number of its vertices, edges, and faces, respectively. Obviously, as Figure 6.1 shows, the numbers $n_v$, $n_e$, and $n_f$ themselves vary wildly from polyhedron to polyhedron, but consider the alternating sum $n_v - n_e + n_f$. One can check that this sum is 2 for all the polyhedra in Figure 6.1. Is this accidental? No, we have just discovered the first combinatorial invariant.

**6.1.1. Theorem** (Euler's Formula). $n_v - n_e + n_f = 2$ for every simple polyhedron.

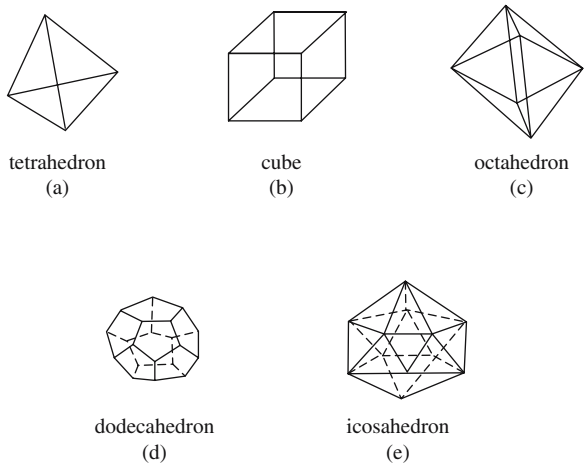



**Figure 6.1.** The five regular polyhedra.

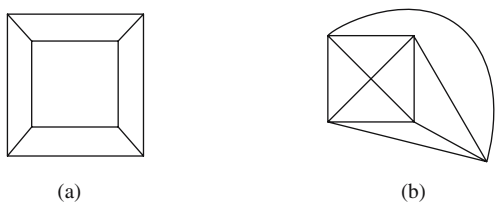



**Figure 6.2.** Flattening the boundary of a cube and octahedron into the plane.

**Sketch of proof.** Since this theorem deals with our first example of an invariant in topology, it is worthwhile showing how this result can be proved. The argument is actually very simple. First, we reduce the problem to a problem in the plane. The boundary of our simple polyhedron is a sphere and so if we remove one face, then the rest of it can be flattened out. Figure 6.2(a) and (b) shows what we would get if we applied this procedure to the cube or octahedron, respectively. We end up with a bounded region **X** in the plane that consists of a collection of $n_v'$ vertices, $n_e'$ edges, and $n_f'$ faces. Clearly,

$$n_v' = n_v, \; n_e' = n_e, \text{ and } n_f' = n_f - 1.$$

The region **X** is homeomorphic to a disk $\mathbf{D}^2$. Our original problem is now equivalent to showing that if a disk has been subdivided into $n_v'$ vertices, $n_e'$ edges (it actually does not matter if the edges are curved or not), and $n_f'$ faces, then

$$n_v' - n_e' - n_f' = 1. \tag{6.1}$$

To simplify our problem, assume that all the faces are triangular. If this is not true already, then we can achieve this by successively adding edges between non-adjacent

vertices of a face. See Figure 6.3. Dividing a face into two does not change the alternating sum $n_v' - n_e' + n_f'$ because we gain both an edge and a face in the process. Next, list the triangular faces in a sequence $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_k$, such that $\mathbf{T}_i$ meets

$$\mathbf{X}_{i-1} = \bigcup_{1 \le j < i} \mathbf{T}_j$$

in either one or two edges. See Figure 6.4. Our argument now proceeds to show equation (6.1) holds for $\mathbf{X}_i$ by induction on i. If i = 1, then $\mathbf{X}_1$ is a triangle and clearly

$$n_v' - n_e' + n_f' = 3 - 3 + 1 = 1.$$

Assume now that equation (6.1) holds for $\mathbf{X}_{i-1}$, $i \ge 2$. As Figure 6.4 shows, when we add $\mathbf{T}_i$ we either increase $n_v'$ and $n_f'$ by 1 and $n_e'$ by 2 (Figure 6.4(a)), or we increase $n_e'$ and $n_f'$ by 1 and leave $n_v'$ unchanged (Figure 6.4(b)). In any case, the sum $n_v' - n_e' - n_f'$ is still equal to 1. This finishes our sketch of the proof of Euler's formula.

The alternating sum $n_v - n_e + n_f$ in Euler's formula is called the *Euler characteristic* of the boundary of the polyhedron and we have just proved that it is a topological invariant. It is the first known result in *combinatorial topology*. The term "combinatorial" is derived from the fact that one is studying invariants based on combinations of numbers, such as $n_v$, $n_e$, and $n_f$, that are easily computed by simple counting.

Euler's formula is a special case of far-reaching generalizations that have many beautiful consequences, some of which we shall see later in this chapter and in the next two chapters. Even this simple version is enough to prove some interesting facts.

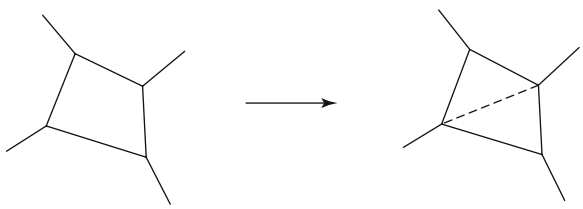


**Figure 6.3.** Subdividing faces into triangles.


**Figure 6.4.** Listing the triangles in a triangular decomposition of a disk.

For excellent papers on the significance and history of Euler's theorem see [Gott96], [GrüS94], [HilP94], and [HilP96]. As an example, we shall now give a simple proof of the following fact that was already known to Euclid.

**6.1.2. Theorem.** There are no regular polyhedra other than the ones shown in Figure 6.1.

**Proof.** Assume that we have a regular polyhedron for which every face has h edges and every vertex belongs to k edges. Since every edge has two vertices and belongs to exactly two faces, it follows that

$$n_f h = 2n_e = n_v k.$$

Substituting these identities into Euler's formula, we get

$$2\frac{n_e}{k} - n_e + 2\frac{n_e}{h} = 2,$$

or equivalently,

$$\frac{1}{n_e} = \frac{1}{h} + \frac{1}{k} - \frac{1}{2}. \tag{6.2}$$

Now h and k are always assumed to be larger than 2 for a polyhedron. On the other hand, if **both** h and k were larger than 3, then equation (6.2) would imply that

$$0 < \frac{1}{n_e} = \frac{1}{h} + \frac{1}{k} - \frac{1}{2} \le \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0,$$

which is impossible. Therefore, either h or k must equal 3. If h = 3, then

$$0 < \frac{1}{n_e} = \frac{1}{3} + \frac{1}{k} - \frac{1}{2}$$

implies that $3 \le k \le 5$. By symmetry, if k = 3, then $3 \le h \le 5$. It follows that

$$(h,k,n_e) = (3,3,6),\ (4,3,12),\ (3,4,12),\ (5,3,30), \text{or } (3,5,30).$$

These values are in fact realized by the tetrahedron, the cube, the octahedron, the dodecahedron, and the icosahedron, respectively. See Figure 6.1 again.

Other early results in topology dealt with the following:

**Map-coloring problems:** The problem is to find the smallest number of colors required to color an arbitrary map in such a way that no two adjacent countries have the same color. For example, the countries A, B, C, and D in Figure 6.5 show that it takes at least four colors. Countries such as E and G that meet in a single point are not considered adjacent. The conjecture that four colors suffice to color any map was finally proved in 1976 with the help of a computer ([AppH77]). The map-coloring

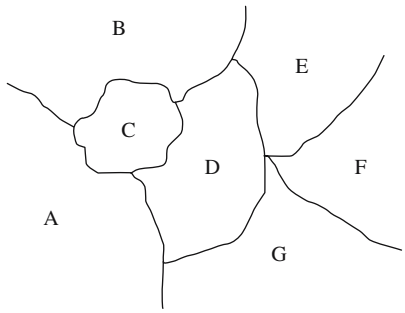



**Figure 6.5.** Coloring maps.

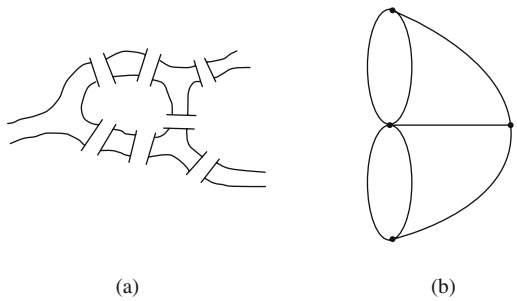



**Figure 6.6.** The Königsberg bridges problem.

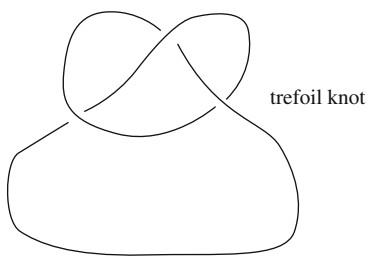



**Figure 6.7.** A nontrivial knot.

problem has been generalized to coloring maps on other surfaces such as on a torus. Interestingly enough, the Euler characteristic, generalized to arbitrary surfaces, shows up in the formula for the smallest number of colors for such maps.

**Graph problems:** A famous example of this type of problem is the Königsberg bridges problem. The problem was to prove that the seven bridges across the Pregel river in the Prussian city of Königsberg could not be crossed by walking without walking across one of the bridges more than once. See Figure 6.6(a). The corresponding graph theory problem, to traverse a graph using each edge exactly one, is shown in Figure 6.6(b).

**Knot theory problems:** Here one wants to classify knots that are thought of as imbeddings of circles in $\mathbf{R}^3$. The unit circle in $\mathbf{R}^2$ is a trivial knot. An example of a nontrivial knot is the *trefoil knot* shown in Figure 6.7. One seeks invariants with which

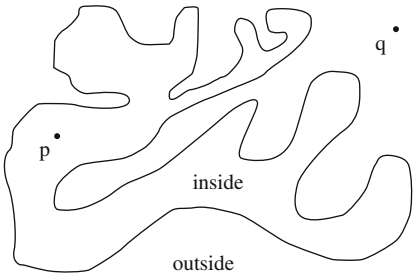



**Figure 6.8.** Does a simple closed curve have an inside and outside?

knots, that is, **imbeddings**, can be distinguished. On an intuitive level, the question is, given a pile of string, what test will determine if it is knotted and when are two knots the same?

**The classification of surfaces:** The problem is to determine computable invariants that will differentiate between different surfaces such as the sphere and the torus.

**The Jordan curve theorem:** A circle in the plane clearly divides the plane into two parts – a bounded part that is the inside of the circle and an unbounded part, the outside. A natural question to ask is whether every simple closed curve in the plane has the same property. Is there again a well-defined inside and outside? See Figure 6.8. Does every curve that starts at an inside point **p** and ends at an outside point **q** have to cross the curve at some point? Surprisingly, this seemingly obvious fact is difficult to prove for general curves. The first partial proof of this fact is due to C. Jordan in 1893.

We shall look at some aspects of combinatorial topology in this chapter. A lot of the material is derived from [AgoM76]. The study of topological invariants is important not just to mathematicians but also to anyone interested in geometric modeling and computer graphics and this chapter will be a warm up for the theory that will be developed in the next chapter. The next section will try to explain in broad terms what topology is all about. Section 6.3 defines simplicial complexes and polyhedra. The latter form the core of the spaces studied by topology. Section 6.4 introduces cutting and pasting. These basic operations in topology are then applied in Section 6.5 to solve the surface classification problem.

## 6.2 What Is Topology?

The topics introduced in the introduction of this chapter may seem like they are all quite separate, but in fact they all have something in common. For example, the shape or size of faces and sides of a simple polyhedron is unimportant in Euler's formula, the shape of a country is unimportant to the coloring of a map, and whether or not a graph is planar does not depend on the length or straightness of its edges. In other words, we were dealing with properties of objects that were invariant under certain deformations. What are the allowable deformations under which the properties stay

invariant? They are more general than motions or isometries. Preserving distance or curvature of curves or surfaces is not important.

As mentioned in Chapter 5, topology is sometimes called "rubber-sheet geometry" because a *topologist* does not distinguish between two subsets in $\mathbf{R}^n$ if, upon pretending that they are made of rubber, he/she can stretch one into the other without tearing or puncturing anything in the process. Compare the difference between the sphere and an ellipsoid and the sphere and the torus. The difference between the former is clearly minor. The ellipsoid is just a slightly elongated or deformed sphere. On the other hand, note that if one cuts along any circle on a sphere, then one will divide the sphere into two parts. The allowable deformations of spaces that we have in mind should preserve this property. Since cutting the torus along a meridian does not separate it, a torus cannot be a deformed sphere, and we shall consider the sphere and torus to be topologically distinct. As another example, consider the sphere and the disk. An obvious difference between the two is that the latter has a boundary whereas the former does not, but there is also a more subtle difference. Our intuition should tell us that there is no way to flatten out a sphere without first puncturing it. To put it another way, a sphere has a "hole" but a disk does not. A torus can also be thought of as having two "holes," but its "holes" are quite different from the one in a sphere.

There is one important observation that needs to be made in the context of rubber-sheet geometry. When we talk about stretching an object, the stretching is not confined to take place in a fixed $\mathbf{R}^n$ but could take place in any higher-dimensional $\mathbf{R}^{n+k}$ for some k. For example, a nontrivial knot in $\mathbf{R}^3$ like the trefoil knot cannot be deformed into the standard circle **within** $\mathbf{R}^3$ (otherwise it would not be knotted), but it can be deformed into the circle by a deformation that takes place in $\mathbf{R}^4$. The point is that the space in which an object happens to be imbedded is unimportant. In fact, it would be better to forget that it is there at all, because one is only interested in the intrinsic properties of the object itself. In analogy with the concept of Platonic forms one should think of equivalent spaces as being merely two different representations of some **single** ideal object. (Of course, what we are talking about now concerns invariant properties of **objects**. We could also ask about invariants for **maps**. The classification of knots has to do with invariants of imbeddings of circles in $\mathbf{R}^3$. This is a whole other story that is postponed to later.)

The transformations that capture the spirit of rubber-sheet geometry are homeomorphisms and the basic problem that the topologist is trying to solve is determining when two spaces are homeomorphic. It would be nice if there were a simple algorithm that would accomplish this. The search for such an algorithm takes us out of general or point set topology and into the domain of algebraic topology and the special case of combinatorial topology. In *algebraic topology* the goal is to associate to each space certain algebraic invariants such that two spaces will be homeomorphic if and only if they have the same invariants. A good simple example of such an invariant is the Euler characteristic that was discussed earlier. However, note that our point of view has changed. Instead of starting with a polyhedron with a given Euler characteristic and discovering other polyhedra with the same Euler characteristic, we are thinking of the Euler characteristic as a number that is the same for all homeomorphic spaces. Returning to the general case, if we had methods for computing our algebraic invariants, then studying a space would reduce to studying its invariants. In this way topological or geometric questions would reduce to problems in algebra (hence the term "algebraic" topology) for which a great deal of theory has already been devel-

oped. The precise nature of the invariants we are talking about cannot be defined in a few words. We will only be scratching the surface of this topic in this chapter and the next. In Chapter 8 we shall describe aspects of *differential topology*. Differential topology builds on algebraic topology but specializes to classification problems for the important class of differentiable manifolds. The differential structure that these manifolds possess enables one to use constructions that are not available for general spaces.

## 6.3 Simplicial Complexes

This section defines the spaces that constitute the core domain to which combinatorial and algebraic topology applies.

**Definition.** A *(finite) simplicial complex* K in $\mathbf{R}^n$ is a finite collection of simplices in $\mathbf{R}^n$ satisfying:

(1) If $\boldsymbol{\sigma} \in$ K, then all faces of $\boldsymbol{\sigma}$ belong to K.
(2) If $\boldsymbol{\sigma}, \boldsymbol{\tau} \in$ K, then either $\boldsymbol{\sigma} \cap \boldsymbol{\tau} = \boldsymbol{\phi}$ or $\boldsymbol{\sigma} \cap \boldsymbol{\tau}$ is a common face of $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$.

**Definition.** Let K be a simplicial complex. The *underlying space* of K, denoted by $|K|$, is defined by

$$|K| = \bigcup_{\boldsymbol{\sigma} \in K} \boldsymbol{\sigma}.$$

The *dimension* of K, denoted by dim K , is defined to be –1 if K is empty and the maximum of the dimensions of the simplices of K, otherwise.

Figure 6.9 shows some examples of simplicial complexes. Note that a simplicial complex is a set of **simplices** and hence **not** a subset of Euclidean space. Its underlying space **is**, however. In practice one is often sloppy with the terminology. In referring to the space in Figure 6.9(c), a person might very well speak of "that simplicial complex K," but as long as the simplices are clearly indicated, there should be no confusion. In the future we may sometimes abbreviate the term "simplicial complex" to "complex."

Condition (1) in the definition of a simplicial complex is a technical one. Its usefulness will become clear later. Condition (2) is the main defining condition and basically states that one should consider a simplicial complex simply as specifying an acceptable decomposition of a space into simplices. This can be done in many ways however, as one can see from Figure 6.10. The two simplicial complexes K and L have the same underlying space.

**Definition.** A simplicial complex L is said to be a *subdivision* of a simplicial complex K if $|K| = |L|$ and every simplex of K is a union of simplices of L.

In Figure 6.10 the simplicial complex L is a subdivision of the simplicial complex K.

A wrong way to subdivide a space into simplices is shown in Figure 6.11(a). The set A is a set of simplices but not a simplicial complex because its two 1-simplices do

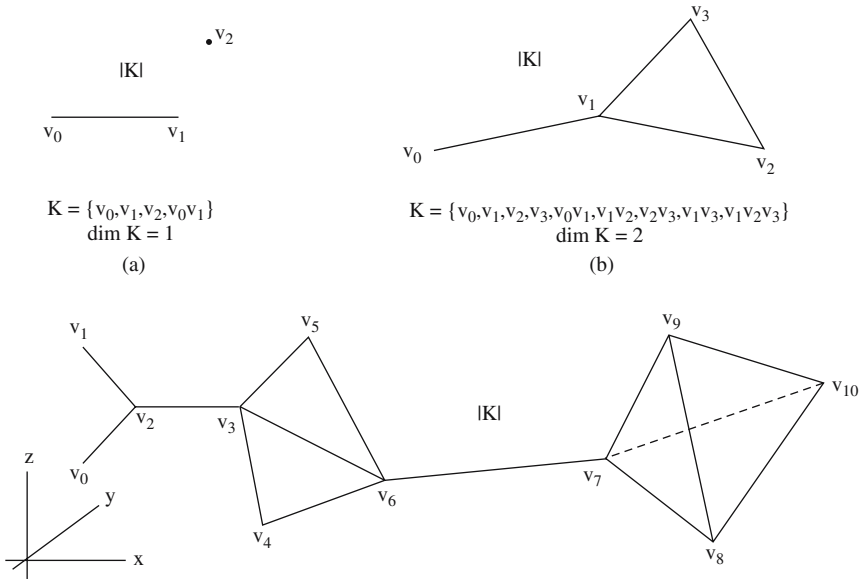



**Figure 6.9.** Examples of simplicial complexes.

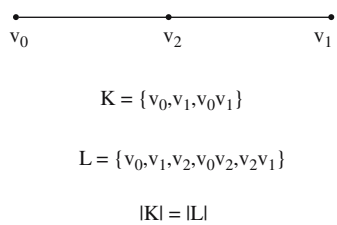



**Figure 6.10.** Different decompositions of an interval.

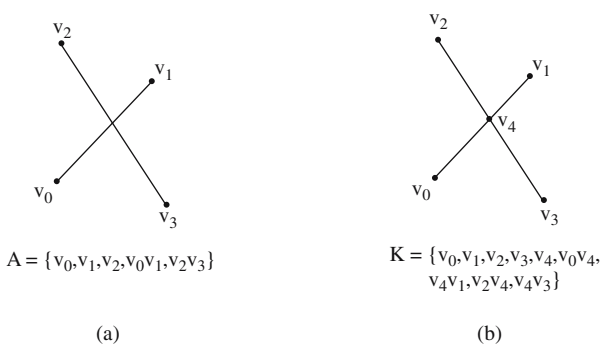



**Figure 6.11.** Invalid and valid simplicial decompositions.

not intersect in a simplex **in** A. On the other hand, $|A|$ is the underlying space of the simplicial complex K in Figure 6.11(b). A useful fact is the following:

**6.3.1. Proposition.** Let K be a simplicial complex. Every point of the underlying space $|K|$ belongs to the interior of a unique simplex of K.

**Proof.** Exercise 6.3.1.

**Definition.** Let K be a simplicial complex. A *subcomplex* of K is a simplicial complex L with $L \subseteq K$.

For example, in Figure 6.10 the set $M = \{\mathbf{v}_0, \mathbf{v}_2, \mathbf{v}_0\mathbf{v}_2\}$ is a subcomplex of L. M is not a subcomplex of K even though $|M| \subseteq |K|$.

**Definition.** The *boundary* of a simplicial complex K, denoted by $\partial K$, is defined by

$$\partial K = \{\boldsymbol{\tau} \mid \boldsymbol{\tau} \text{ is a face of a simplex } \boldsymbol{\sigma}^k \in K \text{ that belongs to a } \textbf{unique} \ (k+1)\text{-simplex of } K\}.$$

It is easy to see that $\partial K$ is a subcomplex of K. For example, if K is the simplicial complex in Figure 6.9(b), then

$$\partial K = \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_1\mathbf{v}_2, \mathbf{v}_2\mathbf{v}_3, \mathbf{v}_1\mathbf{v}_3\}.$$

This example also shows that the underlying space of $\partial K$, $|\partial K|$, is not necessarily the boundary of $|K|$ because, thinking of $|K|$ as a subset of $\mathbf{R}^2$,

$$\text{bdry}\,|K| = |\partial K| \cup \mathbf{v}_0\mathbf{v}_1.$$

We always have $|\partial K| \subseteq \text{bdry}\,|K|$ and the two sets are the same if $|K|$ is an n-dimensional manifold in $\mathbf{R}^n$.

**Definition.** Let K be a simplicial complex and let $\mathbf{v}$ and $\mathbf{w}$ be vertices of K. An *edge path* in K from $\mathbf{v}$ to $\mathbf{w}$ is a sequence of vertices $\mathbf{v} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n = \mathbf{w}$ of K with the property that $\mathbf{v}_i\mathbf{v}_{i+1}$ is a 1-simplex in K for $0 \le i < n$. An edge path from $\mathbf{v}$ to $\mathbf{v}$ is called an *edge loop* at $\mathbf{v}$.

**Definition.** A simplicial complex K is *connected* if, given any two vertices $\mathbf{v}$ and $\mathbf{w}$ in K, there is an edge path from $\mathbf{v}$ to $\mathbf{w}$.

The simplicial complex in Figure 6.9(a) is **not** connected, whereas those in Figures 6.9(b) and 6.9(c) **are**. It is easy to show that a simplicial complex K is connected if and only if $|K|$ is path-connected.

**Definition.** Let $\mathbf{X}$ be a subspace of $\mathbf{R}^n$. A *triangulation* of $\mathbf{X}$ is a pair $(K,\varphi)$, where K is a simplicial complex and $\varphi : |K| \to \mathbf{X}$ is a homeomorphism. The complex K is said to *triangulate* $\mathbf{X}$. A (finite) *polyhedron* is any space that admits a triangulation.

Note that even simple spaces like the unit circle are not the underlying space of a simplicial complex because a "curved" space cannot be built out of "flat" spaces like simplices. It is easy to show, however, that the circle is a polyhedron. In fact, all the spaces that one usually can think of are polyhedra. The only obvious exceptions have simple explanations. For example, the plane is not a polyhedron even though it has a nice triangulation using an "infinite" simplicial complex as shown in Figure 6.12(a). One can easily allow a simplicial complex to have an infinite number of simplices, but when one deals with infinity one does have to be careful. One would not want to think of the reals as a 0-dimensional complex whose simplices are the individual points. That would imply the wrong **topology** for the space. Here is the correct definition:

**Definition.** An *infinite simplicial complex* K is a countably infinite collection of simplices in $\mathbf{R}^n$ satisfying conditions (1) and (2) in the definition of a finite simplicial complex and the following:

(3) Every point of $\mathbf{R}^n$ has a neighborhood that meets only finitely many simplices of K.

An *infinite polyhedron* is a space that is triangulated by an infinite simplicial complex.

All spaces studied in this book are finite or infinite polyhedra. Since we deal mostly with finite ones we shall omit the adjective "finite" and explicitly use the adjective "infinite" on those occasions where we need infinite simplicial complexes or polyhedra. By and large, spaces that "really" are not finite polyhedra are topologically "weird" spaces such as the one shown in Figure 6.12(b) but could nevertheless be useful like fractals in computer graphics (described in Chapter 22 in [AgoM05]).

Sometimes we want to distinguish between "curved" and "flat" or "linear" spaces.

**Definition.** The underlying space of an n-dimensional simplicial complex K is called a *linear polyhedron* of *dimension* n.

It is easy to show that the term "convex linear polyhedron" using the new definition of linear polyhedron is compatible with the term as defined in Section 1.7.

**Definition.** Let $\mathbf{p}_1, \mathbf{p}_2, \ldots,$ and $\mathbf{p}_k$ be a sequence of points in a two-dimensional plane $\mathbf{X}$. The set

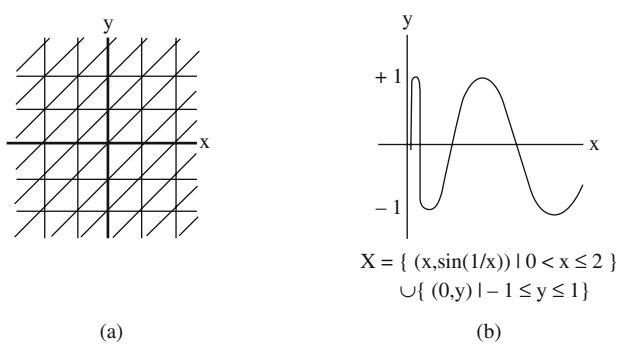



**Figure 6.12.** Spaces without finite triangulations.

$$\mathbf{C} = \bigcup_{i=1}^{k-1} [\mathbf{p}_i, \mathbf{p}_i + 1]$$

is called a *polygonal curve* in $\mathbf{X}$ defined by the vertex sequence. Each $\mathbf{p}_i$ is called a *vertex* of the curve. If $\mathbf{p}_1 = \mathbf{p}_k$, then $\mathbf{C}$ is called a *closed* polygonal curve.

Let $\mathbf{C}$ be a closed polygonal curve in a plane $\mathbf{X}$ in $\mathbf{R}^n$ that is homeomorphic to a circle. By the Jordan curve theorem the curve $\mathbf{C}$ divides the plane into two connected closed sets $\mathbf{A}$ and $\mathbf{B}$, so that

$$\mathbf{X} = \mathbf{A} \cup \mathbf{B}, \quad \mathbf{A} \cap \mathbf{B} = \mathbf{C},$$

and one of the parts, say $\mathbf{A}$, is bounded and the other is not.

**Definition.** The bounded part $\mathbf{A}$ is called the *polygon* with *vertices* $\mathbf{p}_i$ defined by the polygonal curve.

A polygon is actually a polyhedron that is homeomorphic to a disk. This fact, that an imbedding of a circle in the plane extends to an imbedding of the disk, is the so-called Schoenflies theorem that was partially proved by A. Schoenflies in 1906.

**Definition.** If $\mathbf{P}$ is a polygon and if $\mathbf{P}_1, \mathbf{P}_2, \ldots,$ and $\mathbf{P}_k$, $k > 0$, are disjoint polygons contained in the interior of $\mathbf{P}$, then removing the interior of a $\mathbf{P}_i$ creates a *hole* in $\mathbf{P}$ and

$$\mathrm{cls}\,(\mathbf{P} - \mathbf{P}_1 - \mathbf{P}_2 - \ldots - \mathbf{P}_k)$$

is called a *polygon with k holes*.

From the comments above, polygons and polygons with holes are two-dimensional polyhedra.

**Definition.** If $\sigma$ is a simplex, then the simplical complex $\langle\sigma\rangle = \{\tau \mid \tau \prec \sigma\}$ is called the *simplicial complex determined by* $\sigma$.

For example, $\langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 \rangle = \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0\mathbf{v}_1, \mathbf{v}_1\mathbf{v}_2, \mathbf{v}_0\mathbf{v}_2, \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\}$.
Next, let us isolate the maps that are naturally associated to complexes.

**Definition.** Let K and L be simplicial complexes. A *simplicial map* $f : K \to L$ is a map f from the vertices of K to the vertices of L with the property that if $\mathbf{v}_0, \mathbf{v}_1, \ldots,$ and $\mathbf{v}_k$ are the vertices of a simplex of K, then $f(\mathbf{v}_0), f(\mathbf{v}_1), \ldots,$ and $f(\mathbf{v}_k)$ are the vertices of a simplex in L. If f is a bijection between the vertices of K and those of L, then f is called an *isomorphism* between K and L and the complexes are said to be *isomorphic*. We shall use the notation $K \approx L$ to denote that complexes K and L are isomorphic.

**6.3.2. Proposition.** Composites of simplicial maps are again simplicial maps.

**Proof.** Easy.

We show how simplicial maps induce continuous maps of underlying spaces. Let

$$f: K \to L$$

be a simplicial map between simplicial complexes K and L. Define a map

$$|f|: |K| \to |L|$$

as follows: Let $\mathbf{x} \in |K|$. By Proposition 6.3.1, the point $\mathbf{x}$ belongs to the interior of some unique simplex $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k$ of K. If the $t_i$ are the barycentric coordinates of $\mathbf{x}$ with respect to $\boldsymbol{\sigma}$, then

$$\mathbf{x} = \sum_{i=0}^{k} t_i \mathbf{v}_i,$$

and we define

$$|f|(\mathbf{x}) = \sum_{i=0}^{k} t_i f(\mathbf{v}_i).$$

**Definition.** The map $|f|$ is called the map of underlying spaces *induced by the simplicial map* f.

**6.3.3. Proposition.**

(1) $|f|$ is a well-defined continuous map.
(2) $|f|$ is a homeomorphism if and only if f is an isomorphism.
(3) If $f : K \to L$ and $g : L \to M$ are simplicial maps, then $|g \circ f| = |g| \circ |f|$.

**Proof.** Exercise 6.3.6.

## 6.4 Cutting and Pasting

We begin by discussing a slight generalization of simplicial complexes. There are two reasons for introducing the abstract simplicial complexes defined below. One is that simplicial complexes, sometimes called geometric complexes, play only an intermediate role in the study of polyhedra. It is the abstract part of their definition that one typically exploits and not the fact that they happen to correspond to a particular subdivision into simplices of an actual space in $\mathbf{R}^n$. This point will be brought home by various constructions we carry out in this section and the next. A second reason is that in topology one often talks about "cutting" a space or "pasting together" (or "identifying") parts of spaces. It often helps tremendously in understanding complicated spaces and constructions by defining them in terms of such cutting and pasting operations. The mathematical basis of these operations is the concept of a quotient space as defined in the last chapter, but because of the special nature of what we are doing in this chapter, we can define that concept more simply using abstract simplicial complexes. Hopefully, this will also strengthen the reader's intuition about quotient spaces in general.

**Definition.** An *abstract simplicial complex* is a set $\mathsf{A}$ of nonempty subsets of a given set V such that

(1) $\{\mathbf{v}\} \in \mathsf{A}$ for every $\mathbf{v} \in V$, and
(2) if $S \in \mathsf{A}$, then every nonempty subset of S belongs to $\mathsf{A}$.

The elements of $\mathsf{A}$ are called *(abstract) simplices*. If $S \in \mathsf{A}$ and if S has k + 1 elements, then S is called an *(abstract) k-simplex*. The elements of V are called the *vertices* of $\mathsf{A}$ and one identifies the vertex $\mathbf{v}$ in V with the 0-simplex $\{\mathbf{v}\}$ in $\mathsf{A}$.

Every simplicial complex K defines an abstract simplicial complex $\mathsf{A}_K$ in a natural way, namely,

$$\mathsf{A}_K = \{\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k\} \mid \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_k \text{ is a k-simplex of K}\}.$$

It is easy to check that $\mathsf{A}_K$ is in fact an abstract simplicial complex. For example, if L is as in Figure 6.10, then

$$\mathsf{A}_L = \{v_0, v_1, v_2, \{v_0, v_2\}, \{v_1, v_2\}\}.$$

Conversely, it is possible to associate a geometric complex to an abstract simplicial complex.

**Definition.** Let $\mathsf{A}$ be an abstract simplicial complex. A *geometric realization* of $\mathsf{A}$ is a pair $(\varphi, K)$, where K is a simplicial complex and $\varphi$ is a bijective map from the vertices of $\mathsf{A}$ to the vertices of K such that $\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a k-simplex of $\mathsf{A}$ if and only if $\varphi(\mathbf{v}_0)\varphi(\mathbf{v}_1)\ldots\varphi(\mathbf{v}_k)$ is a k-simplex of K.

To show that geometric realizations exist, we simply need to "fill in" the missing points in the abstract simplices.

**6.4.1. Theorem.** Every abstract simplicial complex $\mathsf{A}$ has a unique (up to isomorphism) geometric realization.

**Proof.** Let V be the set of vertices of $\mathsf{A}$ and assume that V has n + 1 points. Let $\varphi$ be a bijection between V and the set of vertices of any n-simplex $\boldsymbol{\sigma}$ in $\mathbf{R}^n$. Define a subcomplex K of $\boldsymbol{\sigma}$ by

$$K = \{\varphi(\mathbf{v}_0)\varphi(\mathbf{v}_1)\cdots\varphi(\mathbf{v}_k) \mid \{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k\} \text{ is a k-simplex of } \mathsf{A}\}.$$

It is easy to see that $(\varphi, K)$ is a geometric realization of $\mathsf{A}$. If $(\varphi', K')$ is another geometric realization of $\mathsf{A}$, then $\varphi \circ \varphi^{-1} : K \to K'$ is an isomorphism.

Theorem 6.4.1 is much more significant than one might conclude from its trivial proof. It is this theorem that allows us to define certain quotient spaces without any fancy point set topology.

Before we show how abstract complexes can be used to make sense of the cutting and pasting operations that we referred to at the beginning of this section, it will help the reader understand what we are talking about here by giving some examples. Prob-

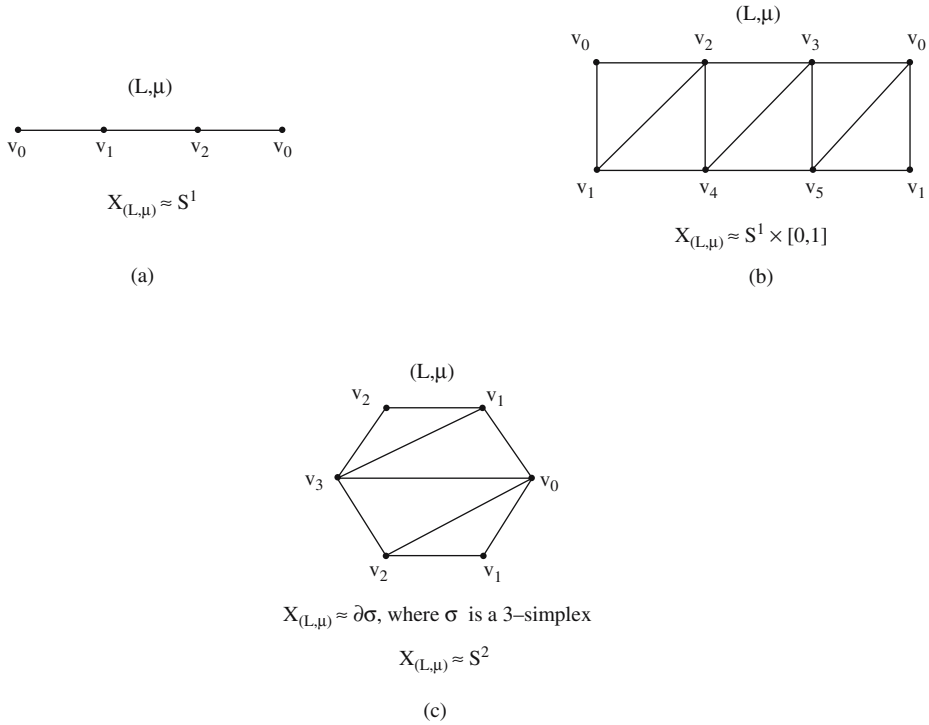



**Figure 6.13.** Spaces as labeled figures.

ably the simplest way to describe a space to someone is to do it visually. In practice, this would mean drawing a picture of it on a flat surface. Consider two-dimensional spaces such as surfaces. Unfortunately, few of even these special spaces can be flattened out into a plane. Most are also too complicated for an easily understood representation by means of projections, but there is another approach. Suppose that we were able to "cut" the space in a few places and that the result could then be flattened out. It turns out that as long as we label the resulting flattened figure appropriately to indicate where the cuts were made, then it will be possible for a person to reconstruct the space mentally from this labeled figure. As an example of how labeled figures can describe spaces, consider Figure 6.13. Does the reader see how Figures 6.13 (a)–(c) are defining a circle, cylinder, and 2-sphere, respectively? Our next task is to make the passage from a labeled figure to the space it represents rigorous. This is where abstract complexes come in.

**Definition.** A *labeled (simplicial) complex* is a triple (L,μ,S), where L is a simplicial complex, S is a set, and μ is a map from the vertices of L to S. The elements of S will be called *labels*. To simplify the notation, we shall usually drop the explicit reference to S and talk about the "labeled complex" (L,μ) whenever S is clear from the context.

Suppose that (L,μ) is a labeled complex. Define an abstract simplicial complex $A_{(L,\mu)}$ by

$$A_{(L,\mu)} = \{\mu(S) \mid S \in A_L\}.$$

Let $(\varphi_{(L,\mu)}, K_{(L,\mu)})$ denote any geometric realization of $A_{(L,\mu)}$ and let $\mathbf{X}_{(L,\mu)} = |K_{(L,\mu)}|$. Let

$$c_{(L,\mu)} : L \to K_{(L,\mu)}$$

be the simplicial map defined on a vertex $\mathbf{v}$ of L by

$$c_{(L,\mu)}(\mathbf{v}) = \varphi_{(L,\mu)}(\mu(\mathbf{v}))$$

and let

$$p_{(L,\mu)} = |c_{(L,\mu)}| : |L| \to \mathbf{X}_{(L,\mu)}.$$

**Definition.** $A_{(L,\mu)}$ is called the abstract simplicial complex *induced* by the labeled complex $(L,\mu)$. $K_{(L,\mu)}$ is called a simplicial complex *defined* by $(L,\mu)$. The space $\mathbf{X}_{(L,\mu)}$ is called a *geometric realization* of $(L,\mu)$ and the map $p_{(L,\mu)}$ is called the *natural projection* of $|L|$ onto $\mathbf{X}_{(L,\mu)}$.

The labeled complex $(L,\mu)$ defines the simplicial complex $K_{(L,\mu)}$ and the space $\mathbf{X}_{(L,\mu)}$ uniquely up to isomorphism and homeomorphism, respectively, with $\mathbf{X}_{(L,\mu)}$ just being a quotient space of $|L|$. It is easy to show that $c_{(L,\mu)}$ and $p_{(L,\mu)}$ are an isomorphism and homeomorphism, respectively, if and only if $\mu$ is a bijection. A good exercise for the reader is to return to Figures 6.13(a)–(c) and show that the spaces $\mathbf{X}_{(L,\mu)}$ are in fact the ones indicated by working through the definitions we have just given. Although converting labeled figures to the spaces they represent is very easy once one understands what is going on, one does have to exercise a little caution. For example, a quick glance at the labeled figure in Figure 6.14(a) might lead one to believe that one is representing a cylinder. This is incorrect. The space $\mathbf{X}_{(L,\mu)}$ is actually homeomorphic to the sphere $\mathbf{S}^2$, just like Figure 6.13(c). There is also a danger in using too

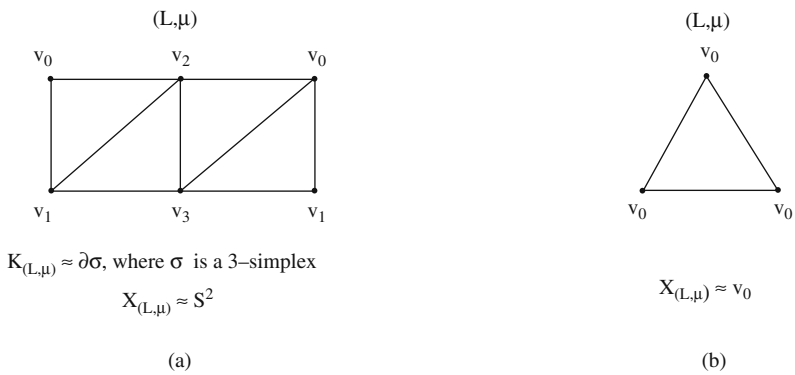



**Figure 6.14.** Why labeled figures have to be interpreted carefully.

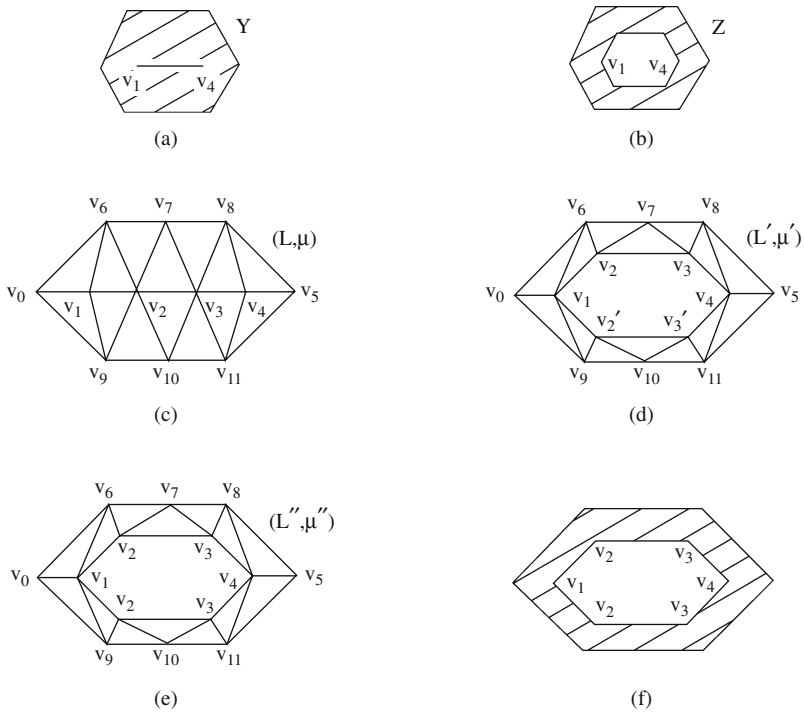



**Figure 6.15.** Using labeled complexes to define cutting and pasting.

few vertices. The labeled complex in Figure 6.14(b) represents a point and not a two-dimensional space.

**Note:** The notation that we have just introduced will be used in other parts of this book. Furthermore, statements such as "the simplicial complex K in Figure A . . . ," where Figure A consists of a labeled complex (L,μ), will always mean "the" complex $K = K_{(L,\mu)}$ (K is well defined up to isomorphism).

Finally, we come to "cutting" and "pasting" and how these operations might be defined. We shall not give a precise definition in general, but now that we know about labeled figures and their geometric realizations, we can at least describe how such a definition might look. Consider Figure 6.15. Intuitively, we can think of the space **Z** in Figure 6.15(b) as the space **Y** in Figure 6.15(a) "cut" along the edge $\mathbf{v}_1\mathbf{v}_4$. Conversely, **Y** is obtained from **Z** by "pasting together" the two arcs from $\mathbf{v}_1$ to $\mathbf{v}_4$ in **Z**. Clearly, from the point of view of point set topology, the space **Y** is nothing but the quotient space of the space **Z** with respect to the equivalence relation on the points of **Z,** which says that points on the two arcs from $\mathbf{v}_1$ to $\mathbf{v}_4$ are related if they are to be glued together. However, by using labeled complexes and their geometric realizations, we can describe the relationship without the formalism of quotient spaces in point set topology.

In the context of the example in Figure 6.15, to define a space **Z** that corresponds to having "cut" **Y** along the edge $\mathbf{v}_1\mathbf{v}_2$, choose a labeled complex $(L,\mu)$ with $|L| = \mathbf{Y}$ as in Figure 6.15(c). If $(L',\mu')$ is the labeled complex shown in Figure 6.15(d), then we can let $\mathbf{Z} = |L'| = \mathbf{X}_{(L',\mu')}$. To define "pasting" we reverse these steps. Assume that we are given $\mathbf{Z} = |L'|$ and $(L',\mu')$ as shown in Figure 6.15(d) and that we want a space **Y** that corresponds to "pasting together" the two arcs from $\mathbf{v}_1$ to $\mathbf{v}_4$ in **Z**. Form the labeled complex $(L'',\mu'')$ as in Figure 6.15(e) by relabeling the vertex $\mathbf{v}_i'$, which is to be identified with the vertex $\mathbf{v}_i$ as $\mathbf{v}_i$ and define $\mathbf{Y} = \mathbf{X}_{(L'',\mu'')}$. Note that $\mathbf{X}_{(L'',\mu'')} \approx \mathbf{X}_{(L,\mu)}$. In practice, one often omits all but the relevant labels in figures. In fact, one may not even specify the triangulation since the homeomorphism type of the resulting space is independent of the choice. For example, one could easily use Figure 6.15(f) to indicate the same type of pasting as Figure 6.15(e).

It should be obvious how to extend the definitions above to the situation where one wants to cut or paste along arbitrary polygonal curves or, more generally, along curves in a polyhedron (use an appropriate triangulation) and therefore feel free to use this terminology in what follows.

## 6.5 The Classification of Surfaces

The classification of surfaces is not only a fascinating chapter in the history of topology but it is also a great place to practice what we learned in the last section. Historically, the earliest topological invariants were discovered in the study of surfaces.

We have already given one definition of surfaces in Chapter 5. They are two-dimensional topological manifolds. For the purposes of this section, we restrict ourselves to **compact** and **connected** surfaces **without boundary** and we now give a combinatorial definition for these

**Definition.** A *combinatorial surface* (without boundary) is a polyhedron **S** together with a triangulation $(K,\varphi)$ satisfying:

(1) K is a two-dimensional connected simplicial complex.
(2) Each 1-simplex of K is a face of precisely two 2-simplices of K.
(3) For every vertex **v** in K, the distinct 2-simplices $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_s$ of K to which **v** belongs can be ordered in such a way that $\boldsymbol{\sigma}_i$, $1 \le i \le s$, meets $\boldsymbol{\sigma}_{i+1}$ in precisely one 1-simplex, where $\boldsymbol{\sigma}_{s+1} = \boldsymbol{\sigma}_1$.

A triangulation $(K,\varphi)$ satisfying properties (1)–(3) is called a *proper triangulation*.

**Note 1.** It is easy to show that every combinatorial surface is in fact a compact connected two-dimensional topological manifold without boundary. The only points where the property of having a neighborhood that is homeomorphic to a disk is not obviously true and needs to be checked are the vertices, but this is where condition (3) is used. The converse is also true:

**6.5.1. Theorem.** (Radó) Every two-dimensional topological manifold (without boundary) in $\mathbf{R}^n$ admits a (possibly infinite) triangulation. If the manifold is compact and connected, then every triangulation is proper.

**Proof.** See [Radó25]. Incidentally, with regard to the first statement in the theorem, if the manifold is compact, then the triangulation will be finite. The infinite triangulations occur with noncompact surfaces such as $\mathbf{R}^2$, which is not a surface using the definition above.

In light of Theorem 6.5.1, we shall drop the adjective "combinatorial" in this chapter and simply refer to a "surface."

The reader may wonder why we have bothered with a rather technical definition of a surface, when we have the more natural manifold definition that corresponds to one's usual intuitive notion about the kind of space that a surface really is. The reason is that our goal is to give a complete classification of surfaces and for this it is convenient to work with proper triangulations. Thus, the technical aspects would not have been avoided. At least with our choice of definition we do not have to appeal to a theorem whose proof would involve a lengthy digression if we were to give it.

**Note 2.** The second part of Theorem 6.5.1 is important because a space can be triangulated in many ways. It would be a very unsatisfactory state of affairs if some triangulations were proper and others not.

**Note 3.** Radó's proof that every two-dimensional topological manifold can be triangulated used methods from complex analysis. The obvious generalization that every topological n-dimensional manifold can be triangulated remained a famous unsolved problem called the *triangulation problem*. We need to point out though that historically when searching for a triangulation for a manifold one was not satisfied with just any triangulation. One assumed a weak regularity condition on the "star" of each vertex.

**Definition.** Let K be a simplicial complex and $\boldsymbol{\sigma}$ a simplex in K. The *star* of $\boldsymbol{\sigma}$, denoted by star($\boldsymbol{\sigma}$), is the union of all the simplices of K that have $\boldsymbol{\sigma}$ as a face, that is,

$$\text{star}\,(\boldsymbol{\sigma}) = \cup\,\{\boldsymbol{\tau} \mid \boldsymbol{\tau} \in \mathrm{K} \text{ and } \boldsymbol{\sigma} \prec \boldsymbol{\tau}\}\,.$$

**Definition.** Two simplicial complexes are said to be *combinatorially equivalent* if they have isomorphic subdivisions.

**Definition.** Call a triangulation (K,φ) for a topological manifold a *proper triangulation* if the subcomplexes that triangulate the boundary of the stars of vertices are combinatorially equivalent to the boundary simplicial complex of an n-simplex. A topological manifold that admits a proper triangulation is called a *combinatorial manifold*.

The condition for a proper triangulation is stronger than just saying that the stars are homeomorphic to $\mathbf{D}^n$ and is basically a generalization of the proper triangulations defined above for surfaces. In 1952 Moise [Mois52] proved that all topological three-dimensional manifolds could be triangulated. One of the major achievements of the 1960s was the solution to the triangulation problem by Kirby and Siebenmann ([KirS69]). They proved that in each dimension n, $n \geq 5$, there are topological mani-

folds that are **not** combinatorial manifolds. Subsequently, Andrew Casson proved in 1985 that there exist four-dimensional topological manifolds that could not be triangulated by **any** simplicial complex. There is another famous related problem, originally a conjecture, called the *Hauptvermutung*. This problem, which was basically a question about the uniqueness of triangulations, asked if any two proper triangulations of a manifold had isomorphic subdivisions. Here also the answer is no. In dimensions larger than four, there are well-defined invariants that determine whether the answers to the above questions are yes or no.

**Note 4.** A space that is a surface using the definition above is also called a *closed* surface. The adjective "closed" in the context of surfaces or manifolds means that there is no boundary. Section 6.6 will briefly discuss the definition and classification of more general types of surfaces, such as surfaces with boundary like the unit disk and the torus with two open disks removed (see Figure 6.16), and noncompact surfaces with or without boundary, such as the open unit disk or that disk with an open disk removed from its interior. We also repeat for emphasis that in this section, unless stated otherwise, a space that is called a "surface" is a compact and connected space.

After these preliminary remarks about surfaces we start the main task of this section, which is their classification. We would like a list of all possible surfaces (up to homeomorphism) and, if possible, a simple characterization of each. The basic step will involve taking a given arbitrary surface, cutting it into pieces, and then reassembling the pieces into some recognizable form. As we cut we will use labels to remember that the two edges that are created are actually identified in the surface. Our proof of the classification theorems will therefore involve the manipulation of lots of labeled complexes. This is why we discussed such complexes in the previous section. In our case here the complexes will actually be planar and the labeling will be modified somewhat to simplify things.

As an example, consider the torus in Figure 6.17(a). Cutting along the circle $\mathbf{A}_1$ produces Figure 6.17(b). Next, cutting along the edge $\mathbf{A}_2$ in Figure 6.17(b), or along

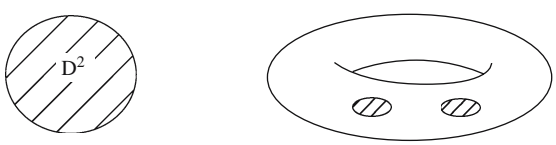



**Figure 6.16.** Surfaces with boundary.

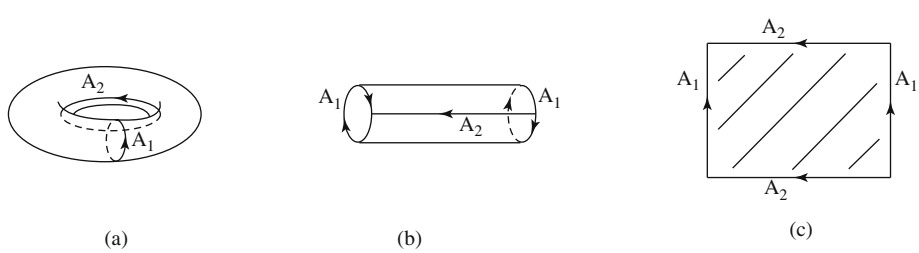



**Figure 6.17.** Cutting and pasting for the torus.

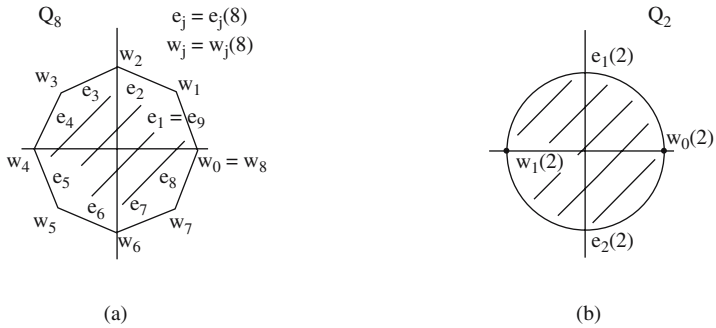



**Figure 6.18.** Regular k-gons.

the circles $\mathbf{A}_1$ and $\mathbf{A}_2$ in Figure 6.17(a) simultaneously, produces Figure 6.17(c). Clearly, Figures 6.17(b) and (c) can be considered representations of a torus if we understand the labeling and arrows properly. Notice that we are now labeling the edges and not the vertices. This will be more intuitive for the cutting and pasting we want to do. It is easy to pass between the two types of labeling however. The main advantage to labeling vertices is that the results about geometric realizations of labeled complexes are easier to state and prove since it is more straightforward to relate this labeling to abstract complexes. At any rate, our first step in classifying surfaces will be to show that an arbitrary surface $\mathbf{S}$ can be represented by a labeled polygon similar to the one we got for the torus. We will also show that this geometric presentation is equivalent to an "algebraic" presentation that consists of a formal symbol.

Let $k \geq 3$ and let $\mathbf{Q}_k$ denote the "standard" regular k-gon (k-sided polygon), namely, the convex hull of the points

$$\mathbf{w}_j(k) = (\cos 2\pi j/k, \sin 2\pi j/k) \in \mathbf{S}^1 .$$

Let

$$\mathbf{e}_j(k) = \mathbf{w}_{j-1}(k)\mathbf{w}_j(k)$$

denote the jth edge of $\mathbf{Q}_k$. Figure 6.18(a) shows $\mathbf{Q}_8$. Since it will also be convenient to have a two-sided "polygon," let $\mathbf{Q}_2 = \mathbf{D}^2$ be the "polygon" with vertices

$$\mathbf{w}_0(2) = (1,0) \quad \text{and} \quad \mathbf{w}_1(2) = (-1,0)$$

and "edges"

$$\mathbf{e}_1(2) = \mathbf{S}^1_+ \quad \text{and} \quad \mathbf{e}_2(2) = \mathbf{S}^1_- .$$

See Figure 6.18(b).

The next lemma basically proves that any surface can be flattened out into the plane by cutting it appropriately. More precisely, it shows that a surface can be rep-

resented as a labeled planar polygon and that we can normalize that polygon to be a regular k-gon.

**6.5.2. Lemma.** Given a surface $\mathbf{S}$ we can always find a labeled complex $(L,\mu) = (L_{\mathbf{S}},\mu_{\mathbf{S}})$ satisfying:

(1) $|L| = \mathbf{Q}_k$ for some k.
(2) The vertices of L are precisely the points $\mathbf{w}_0(k)$, $\mathbf{w}_1(k)$, . . . , and $\mathbf{w}_{k-1}(k)$.
(3) There is a homeomorphism $h : \mathbf{X}_{(L,\mu)} \to \mathbf{S}$.
(4) If $\psi_{\mathbf{S}} : \mathbf{Q}_k \to \mathbf{S}$ is given by $\psi_{\mathbf{S}} = h \circ p_{(L,\mu)}$, then $\psi_{\mathbf{S}}|\mathrm{int}(\mathbf{Q}_k)$ and $\psi_{\mathbf{S}}|\mathbf{e}_i(k)$ are one-to-one.

**Proof.** See [AgoM76].

Think of the labeled complex $(L_{\mathbf{S}},\mu_{\mathbf{S}})$ in Lemma 6.5.2 as being derived from $\mathbf{S}$ by cutting along $\psi_{\mathbf{S}}(\partial\mathbf{Q}_k)$. Conversely, $\mathbf{S}$ can be reconstructed from $\mathbf{Q}_k$ by pasting together those edges of $\mathbf{Q}_k$ that are mapped onto the same set in $\mathbf{S}$ by $\psi_{\mathbf{S}}$.

The main consequence of Lemma 6.5.2 is that the study of surfaces has been reduced to the study of certain labeled complexes because each surface $\mathbf{S}$ has an associated labeled complex $(L_{\mathbf{S}},\mu_{\mathbf{S}})$, which in turn determines the surface since $\mathbf{S} \approx \mathbf{X}_{(L_{\mathbf{S}},\mu_{\mathbf{S}})}$.

Before describing an even simpler and more compact representation for $\mathbf{S}$ we need some notation. Let

$$\Sigma_+ = \{A_1, A_2, \ldots\}$$

be the infinite set of distinct symbols $A_i$. Define

$$\Sigma = \{A_1, A_2, \ldots\} \cup \{A_1^{-1}, A_2^{-1}, \ldots\},$$

where each expression $A_i^{-1}$ is considered as a purely formal symbol and no algebraic significance is attached to the superscript "–1." We shall identify the symbol $(A_i^{-1})^{-1}$ with $A_i$. With this identification $a^{-1}$ will belong to $\Sigma$ whenever the symbol a does. Let

$$\Omega = \text{the set of all nonempty finite strings } a_1a_2 \ldots a_q, \text{ where } a_i \in \Sigma .$$

For example, the strings

$$A_1A_1 = A_1(A_1^{-1})^{-1}, \quad A_1A_2A_1^{-1}A_2^{-1}, \quad \text{and} \quad A_1A_1A_1A_1^{-1}A_1A_3$$

belong to $\Omega$.

Returning to our surface $\mathbf{S}$, choose a labeled complex $(L_{\mathbf{S}},\mu_{\mathbf{S}})$ and map $\psi_{\mathbf{S}} : \mathbf{Q}_k \to \mathbf{S}$ as described in Lemma 6.5.2. Given an "admissible" labeling of the edges of $\mathbf{Q}_k$, define a string

$$w_{\mathbf{S}} = a_1a_2 \ldots a_k \in \Omega$$

for $\mathbf{S}$ by letting the element $a_i$ be the label of the ith edge $\mathbf{e}_1(k)$ of $\mathbf{Q}_k$. Here is an informal description of how such admissible labelings are obtained. Label the first edge

$\mathbf{e}_1(k)$ of $\mathbf{Q}_k$ with an arbitrary element from $\Sigma$. Continue around to the other edges of $\mathbf{Q}_k$ in a counterclockwise fashion and associate a different label from $\Sigma$ to each of them (if "a" has been used, "$a^{-1}$" does not count as different) **unless** the new edge, say $\mathbf{e}_i = \mathbf{e}_i(k)$, is identified with a previously labeled edge, say $\mathbf{e}_j = \mathbf{e}_j(k)$. In that case, the labels for $\mathbf{e}_i$ and $\mathbf{e}_j$ should reflect this identification while at the same time distinguishing between the two possible ways that the edges could be identified. Assume that $\mathbf{e}_j$ has been labeled "a." If the edges are identified in an orientation-preserving way, then use the label "a" for $\mathbf{e}_i$, otherwise use "$a^{-1}$."

A more rigorous recursive definition of the string $w_{\mathbf{S}}$ is the following. Define $a_1$ to be an arbitrary element of $\Sigma$. Let $2 \le i \le k$ and assume that the elements $a_1, a_2, \ldots,$ and $a_{i-1}$ have already been defined. The definition of $a_i$ divides into two cases:

**Case 1.** $\psi_{\mathbf{S}}(\mathbf{e}_i(k)) \ne \psi_{\mathbf{S}}(\mathbf{e}_j(k))$ for all j, $1 \le j < i$ : In this case, let $a_i$ be an arbitrary element of

$$\Sigma - \{a_1, a_1^{-1}, a_2, a_2^{-1}, \ldots, a_{i-1}^{-1}\}$$

**Case 2.** $\psi_{\mathbf{S}}(\mathbf{e}_i(k)) = \psi_{\mathbf{S}}(\mathbf{e}_j(k))$ for some j, $1 \le j < i$ : In this case, let

$$\begin{aligned} a_i &= a_j \quad \text{if} \quad \psi_{\mathbf{S}}(\mathbf{w}_{i-1}(k)) = \psi_{\mathbf{S}}(\mathbf{w}_{j-1}(k)) \\ &= a_j^{-1} \quad \text{otherwise.} \end{aligned}$$

**Definition.** The string $w_{\mathbf{S}} \in \Omega$ is called a *symbol associated to* $(L_{\mathbf{S}}, \mu_{\mathbf{S}})$, or simply a *symbol* for the surface $\mathbf{S}$.

Note that the symbol for a surface is not unique since there is no unique choice of $a_i$ in Case 1 above.

**6.5.3. Example.** To find a symbol for the sphere $\mathbf{S}^2$.

**Solution.** Suppose that we have triangulated $\mathbf{S}^2$ with the complex

$$K = \partial\langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3 \rangle ,$$

where $\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$ is some 3-simplex. The first task is to find a labeled complex for some regular k-gon of the type guaranteed by Lemma 6.5.2. Although this would not be hard to do directly in our special case, in general it would be easiest to use two steps: One would first flatten the surface out into the plane by cutting and then move the result to a regular k-gon. We shall follow this general approach, which is actually how Lemma 6.5.2 would be proved. See Figure 6.19(a). The first step would produce the simplicial complex L with the simplicial map

$$\alpha : L \to K$$

defined by the condition that

$$\alpha \,|\, \langle \boldsymbol{\sigma}_i' \rangle : \langle \boldsymbol{\sigma}_i' \rangle \to \langle \boldsymbol{\sigma}_i \rangle$$

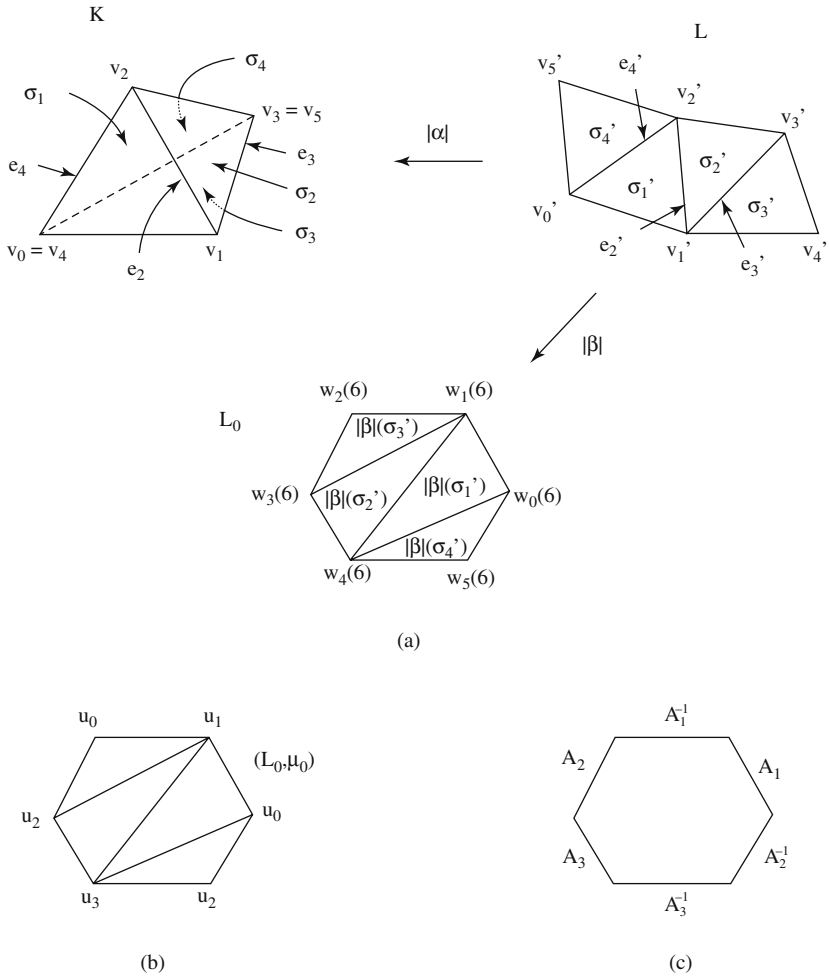



**Figure 6.19.** From labeled complex to symbol for sphere.

is an isomorphism for all i. Think of $|L|$ as having been obtained from $|K|$ by cutting along the edges $\mathbf{v}_0\mathbf{v}_3$ and $\mathbf{v}_0\mathbf{v}_1$. Mapping $|L|$ to a regular k-gon is easy. For example, the triangulation $L_0$ of $\mathbf{Q}_6$ and the simplicial map

$$\beta : L \to L_0$$

$$\beta(\mathbf{v}_i') = \mathbf{w}_i(6)$$

shown in Figure 6.19(a) does the job. It is easy to see that the labeled complex $(L_0,\mu_0)$ shown in Figure 6.19(b) defines our sphere in the sense that a geometric realization of it would be homeomorphic to $\mathbf{S}^2$. $(L_0,\mu_0)$ is a labeled complex of the type we were looking for and we can use the algorithm described above on it to get a symbol such as

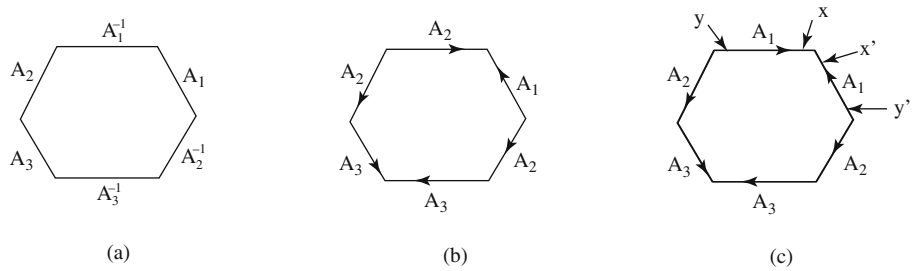



**Figure 6.20.** Alternate labeled complexes for sphere.

$$A_1A_1^{-1}A_2A_3A_3^{-1}A_2^{-1}$$

for $\mathbf{S}^2$. Figure 6.19(c) shows the corresponding labeled polygon.

Example 6.5.3 listed one symbol for a sphere, but it is clear that a different choice of edge labels could have produced the symbol

$$A_4^{-1}A_4A_1A_6A_6^{-1}A_1^{-1}\ .$$

As we stated earlier, symbols for surfaces are not unique. On the other hand, although the labeled complex $(L_0,\mu_0)$ does not determine a unique symbol for the sphere, we leave it to the reader to convince him/herself that all symbols derived from $(L_0,\mu_0)$ will have the form

$$a_1a_1^{-1}a_2a_3a_3^{-1}a_2^{-1}$$

for $a_i \in \Sigma$. In general, one can show that every symbol associated to a labeled complex $(L_{\mathbf{S}},\mu_{\mathbf{S}})$ from Lemma 6.5.2 for a surface $\mathbf{S}$ has the same basic structure, that is, if $a_1a_2 \ldots a_k$ and $b_1b_2 \ldots b_k$ are two symbols associated to $(L_{\mathbf{S}},\mu_{\mathbf{S}})$, then there is a permutation $\sigma$ of $\Sigma$ such that $b_i = \sigma(a_i)$ and $\sigma(a^{-1}) = \sigma(a)^{-1}$ for all $a \in \Sigma$. This justifies our talking about "the" symbol $w_{\mathbf{S}}$ associated to $(L_{\mathbf{S}},\mu_{\mathbf{S}})$ after all.

Continuing our sphere example, even though Figure 6.20(a) is a good pictorial representation for one of its symbols, it is not the one that is usually adopted. By drawing little arrows in the edges of $\mathbf{Q}_6$ as indicated in Figure 6.20(b), one can incorporate, without superscripts on the symbols, the same information that was contained in Figure 6.20(a). The two possible ways of identifying edges (via linear maps) are specified by the direction of the arrows. For example, the arrows tell us that the points $\mathbf{x}$ and $\mathbf{y}$ in Figure 6.20(c) are to be identified with the points $\mathbf{x}'$ and $\mathbf{y}'$, respectively. Observe that the direction of the arrows is not uniquely specified by a symbol. Simultaneously reversing their direction on two edges that are to be identified changes nothing. The only important property that is an invariant is whether these arrows are both in the same or opposite direction. Most of the labeled polygons we shall refer to from now on will have the arrows in their sides rather than superscripts on the labels, but we should remember that either or both methods simultaneously is

permissible. Furthermore, because it is clearly easy to pass back and forth between symbols and labeled polygons $\mathbf{Q}_k$, we can use either of these two representations interchangeably.

We have defined symbols for surfaces, but that is only half the story. We basically have a "map" from the set of surfaces to the set of strings $\Omega$. To really take advantage of this correspondence we need a map that goes the other way, that is, we need a map that associates a surface to a string in $\Omega$. It is clear how to define this map on an intuitive level. Rather than cutting we shall paste. Basically, if $w = a_1a_2 \ldots a_k \in \Omega$, then we start with $\mathbf{Q}_k$ and construct a space $\mathbf{S}_w$ by pasting together any edges $\mathbf{e}_i(k)$ and $\mathbf{e}_j(k)$ of $\mathbf{Q}_k$ whenever $a_i = a_j$ or $a_j^{-1}$. It is easy to see that if w is an arbitrary string, then the space $\mathbf{S}_w$ need not be a surface. Therefore, in addition to explaining the construction of $\mathbf{S}_w$ more carefully, one would also like to know under what conditions $\mathbf{S}_w$ will be a surface.

**Definition.** If $a \in \Sigma$, then define $n_w(a)$ to be the number of times that the symbol a or $a^{-1}$ appears in the string w. The *length of w*, $l(w)$, is defined by

$$l(w) = \sum_{a \in \Sigma} n_w(a) .$$

For example, if $w = A_1A_2A_1^{-1}A_1A_2A_3^{-1}$, then $n_w(A_1) = 3$, $n_w(A_2) = 2$, $n_w(A_3) = 1$, and $n_w(A_i) = 0$ for $i > 3$. Also, $l(w) = 6$.

**Definition.** Define a subset $\Omega^*$ of $\Omega$ by

$$\Omega^* = \{w \in \Omega \mid n_w(a) = 0 \text{ or } 2 \text{ for all } a \in \Sigma\} .$$

It is easy to see that if $w_{\mathbf{S}}$ is a symbol for a surface $\mathbf{S}$, then $w_{\mathbf{S}} \in \Omega^*$. In fact, the next lemma shows among other things that $\mathbf{S}_w$ is a surface if and only if $w \in \Omega^*$.

**6.5.4. Lemma.** There is a construction that associates to each $w \in \Omega^*$ a well-defined labeled complex $(L_w, \mu_w)$ with the following properties:

(1) $|L_w| = \mathbf{Q}_{l(w)}$ if $l(w) > 2$ and $|L_w| = \mathbf{Q}_6$ if $l(w) = 2$.
(2) The space $\mathbf{S}_w = X_{(L_w,\mu_w)}$ is a surface.
(3) If $\mathbf{S}$ is a surface and if u is any symbol for $\mathbf{S}$, then $\mathbf{S} \approx \mathbf{S}_u$.
(4) Let $a, b \in \Sigma$. If $w = aa^{-1}$, then $\mathbf{S}_w \approx \mathbf{S}^2$. If $w = aa$, then $\mathbf{S}_w \approx \mathbf{P}^2$. If $w = aba^{-1}b^{-1}$, then $\mathbf{S}_w \approx \mathbf{S}^1 \times \mathbf{S}^1$.

**Proof.** See [AgoM76]. It is easy to justify part (4). Cutting a sphere along an arc allows us to flatten the remainder into a disk with two edges that are appropriately identified. For $\mathbf{P}^2$, recall the discussion in Section 3.4 and Figure 3.9. For $\mathbf{S}^1 \times \mathbf{S}^1$, see Figure 6.17.

**Definition.** If $\mathbf{S}$ is a surface, then any $w \in \Omega^*$ such that $\mathbf{S}_w \approx \mathbf{S}$ will be called a *symbol for* $\mathbf{S}$.

The fact that this new definition of a symbol for a surface is compatible with the earlier one follows from Lemma 6.5.4(3).

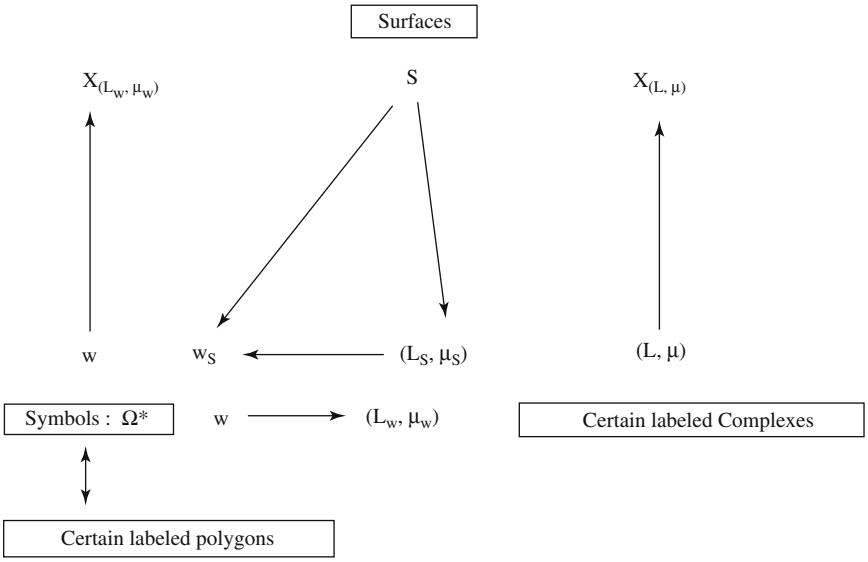



**Figure 6.21.** The relationship between surfaces and labeled objects.

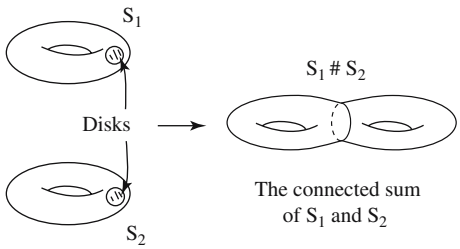



**Figure 6.22.** A connected sum example.

The first step to the classification of surfaces is now completed. Figure 6.21 summarizes the various correspondences that we have established. We have associated to every surface a collection of symbols. Each of these symbols determines the surface uniquely up to homeomorphism, so that whenever two surfaces have a symbol in common they are homeomorphic. Before we move onto the next step we shall find symbols for some more standard surfaces.

We already know from Lemma 6.5.4(4) that

$$A_1A_1^{-1},\ A_1A_1,\ \text{and}\ A_1A_2A_1^{-1}A_2^{-1}$$

are symbols for $\mathbf{S}^2$, $\mathbf{P}^2$, and $\mathbf{S}^1 \times \mathbf{S}^1$, respectively. Knowing this one can determine symbols for other surfaces that can be formed from these basic ones by means of what is called a "connected sum" operation. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two surfaces. Intuitively, the connected sum of $\mathbf{S}_1$ and $\mathbf{S}_2$ is the surface one gets by cutting out a disk from both $\mathbf{S}_1$ and $\mathbf{S}_2$ and pasting the remainders together along the boundaries of the holes that were generated. See Figure 6.22. To make this rigorous, choose a proper triangula-

tion $(K_i,\varphi_i)$ for $\mathbf{S}_i$ and let $\boldsymbol{\sigma}_i$ be a 2-simplex of $K_i$. If the $K_i$ are simplicial complexes in $\mathbf{R}^n$, we can choose n to be large enough so that we can reposition the polyhedra $|K_i|$ in such a way that $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2 = |K_1| \cap |K_2|$.

**Definition.** Any space that is homeomorphic to the underlying space of the simplicial complex

$$L = (K_1 \cup K_2) - \{\boldsymbol{\sigma}_i\}$$

is defined to be the *connected sum of* $\boldsymbol{S_1}$ *and* $\boldsymbol{S_2}$ and is denoted by $\mathbf{S}_1 \# \mathbf{S}_2$.

**6.5.5. Proposition.** The connected sum operation # is well defined, that is, the homeomorphism type of $\mathbf{S}_1 \# \mathbf{S}_2$ does not depend on the triangulations $(K_i,\varphi_i)$ or the choice of simplices $\boldsymbol{\sigma}_i$. Furthermore,

(1) $\mathbf{S}_1 \# \mathbf{S}_2$ is a surface.
(2) (Commutativity) $\mathbf{S}_1 \# \mathbf{S}_2 \approx \mathbf{S}_2 \# \mathbf{S}_1$.
(3) (Associativity) $\mathbf{S}_1 \# (\mathbf{S}_2 \# \mathbf{S}_3) \approx (\mathbf{S}_1 \# \mathbf{S}_2) \# \mathbf{S}_3$.
(4) (Identity) $\mathbf{S}_1 \# \mathbf{S}^2 \approx \mathbf{S}_1$.

**Proof.** The hard part of this proof is to show that the operation # is well defined. The rest is easy. See [AgoM76].

**6.5.6. Proposition.** Table 6.5.1 shows the symbols for the sphere and the connected sum of n tori and n projective planes.

**Proof.** See [AgoM76].

**Definition.** The symbols in Table 6.5.1 are called the *normal forms* for the corresponding surfaces.

The next step in classifying surfaces involves manipulating labeled polygons. One ends up with the next theorem, the first main theorem dealing with the classification of surfaces. On an intuitive level the proof involves starting with a labeled polygon that corresponds to the symbol of a given surface and then, by cutting and pasting, changing that polygon into another labeled polygon that defines a symbol having one of the normal forms shown in Table 6.5.1. The details are lengthy and messy but not hard.

**Table 6.5.1** Surface symbols.

| Surface | Symbol |
|---|---|
| $S^2$ | $aa^{-1}$ |
| $\underbrace{(S^1\times S^1)\#\cdots\#(S^1\times S^1)}_{n\text{ times}}$ | $a_1a_2a_1^{-1}a_2^{-1}\cdots a_{2n-1}a_{2n}a_{2n-1}^{-1}a_{2n}^{-1}$ |
| $\overbrace{P^2\#\cdots\#P^2}$ | $a_1a_1\cdots a_na_n$ |

**6.5.7. Theorem.** Every surface is homeomorphic either to the sphere, or to a connected sum of tori, or to a connected sum of projective planes.

**Proof.** See [AgoM76].

To finish the classification of surfaces we need to show that the surfaces in Theorem 6.5.7 are nonhomeomorphic.

**Definition.** Let **S** be a surface and let $(K,\varphi)$ be any proper triangulation of **S**. Let $n_i(K)$ denote the number of i-simplices in K. The *Euler characteristic of* ***S****, χ(****S****),* is defined by

$$\chi(S) = n_0(K) - n_1(K) + n_2(K).$$

**6.5.8. Proposition.** The Euler characteristic for a surface is a well-defined integer that does not depend on the triangulation of the surface that is chosen for the definition. Furthermore,

(1) $\chi(\mathbf{S}^2) = 2$ , $\chi(\mathbf{S}^1 \times \mathbf{S}^1) = 0$, and $\chi(\mathbf{P}^2) = 1$.
(2) If $\mathbf{S}_1$ and $\mathbf{S}_2$ are surfaces, then $\chi(\mathbf{S}_1 \# \mathbf{S}_2) = \chi(\mathbf{S}_1) + \chi(\mathbf{S}_2) - 2$.

**Proof.** The fact that the Euler characteristic is well defined is a special case of a much more general result proved later in Section 7.4. Part (1) is easily verified from triangulations of the spaces in question. (Any reader who has trouble finding a triangulation for the torus or projective plane can find one in Section 7.2.) To prove (2), let $K_1$, $K_2$, L, $\sigma_1$, and $\sigma_2$ be as in the definition of the connected sum. The complexes $K_i$ and L triangulate the surfaces $\mathbf{S}_i$ and $\mathbf{S}_1 \# \mathbf{S}_2$, respectively, and the 2-simplex $\sigma_i$ belongs to $K_i$. Since L has all the 2-simplices of $K_1$ and $K_2$ except for $\sigma_1$ and $\sigma_2$, it follows that

$$n_2(L) = n_2(K_1) + n_2(K_2) - 2.$$

But, in L, the boundary of $\sigma_1$ has been identified to the boundary of $\sigma_2$. Therefore,

$$n_1(L) = n_1(K_1) + n_1(K_2) - 3$$

and

$$n_0(L) = n_0(K_1) + n_0(K_2) - 3,$$

because one does not want to count the 0- and 1-simplices in the boundary of $\sigma_1$ and $\sigma_2$ twice. The three equations easily lead to the result in part (2) of the proposition.

It follows from Proposition 6.5.8 and Theorem 6.5.7 that it is easy to compute the Euler characteristic of any surface.

We are almost ready to prove the second part of the classification theorem for surfaces. Before we do, we need to bring up orientability again and also define a commonly used term in connection with surfaces, namely the "genus." First of all, the

sphere and the connected sum of tori are orientable surfaces whereas the connected sum of projective planes is not. We shall not be able to prove this, however, until the next chapter and after we have a precise definition of orientability.

**Definition.** If **S** is a surface, define the *genus* of **S** to equal 0 if **S** is homeomorphic to the sphere and equal to n if **S** is homeomorphic to a connected sum of n tori or n projective planes, $n \geq 1$.

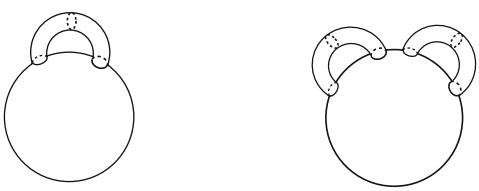
Intuitively, if the genus of an orientable surface is n, then the surface is homeomorphic to a *sphere with n handles*. See Figure 6.23. Since the projective plane is the union of a disk and a Moebius strip, it is often referred to as a sphere with a *crosscap* in the literature. With this terminology, a nonorientable surface of genus n is called a *sphere with n crosscaps*.

A simple formula relates the genus g of a surface **S** to its Euler characteristic $\chi$:

$$\begin{aligned} g &= (2-\chi)/2, && \text{if } \mathbf{S} \text{ is orientable} \\ &= 2-\chi, && \text{otherwise.} \end{aligned} \tag{6.3}$$

The next proposition summarizes what we just shown.

**6.5.9. Proposition.** Table 6.5.2 shows the Euler characteristic, orientability, and genus of the listed surfaces.



$S^1 \times S^1$

Genus 1

$(S^1 \times S^1) \# (S^1 \times S^1)$

Genus 2

**Figure 6.23.** Two surfaces as spheres with handles.

**Table 6.5.2** The geometric invariants of some surfaces.

| Surface | Euler characteristic | Orientability | Genus |
|---|---|---|---|
| $\mathbf{S}^2$ | 2 | orientable | 0 |
| $\underbrace{(S^1\times S^1)\#\cdots\#(S^1\times S^1)}_{\text{n times}}$ | 2 − 2n | orientable | n |
| $\overbrace{P^2\#\cdots\#P^2}$ | 2 − n | nonorientable | n |

**6.5.10. Theorem.** The sphere, the connected sum of n tori, $n \geq 1$, and the connected sum of m projective planes, $m \geq 1$, are nonhomeomorphic spaces.

**Proof.** The proof follows easily from Proposition 6.5.9 and the topological invariance of the Euler characteristic.

Theorems 6.5.7 and 6.5.10 together comprise what is called the classification theorem for closed surfaces. This theorem can be summarized by the following:

**Algorithm for determining the homeomorphism type of a given surface S:**

**Step 1:** Decide whether or not **S** is orientable.
**Step 2:** Compute the Euler characteristic of **S** from a triangulation.
**Step 3:** Look up the surface type in Table 6.5.2.

Computing the Euler characteristic is a simple counting procedure but determining orientability is a little more subtle. The problem is that we do not yet have a formal definition of orientability. We shall return to that point in the next chapter (Section 7.5).

If we do not count connected sums, then the only concrete example of a nonorientable surface that we have so far is the projective plane. There is another well-known nonorientable surface, called the Klein bottle, named after the mathematician Felix Klein who first described it. This surface is easier to visualize than the projective plane and we finish this section with a brief discussion of it.

First, recall that to construct a torus one can start with a tube (Figure 6.17(b)) and then bring the ends around and glue them together (Figure 6.17(a)). We vary this construction slightly. We again start with a tube (Figure 6.24(a)), but rather than

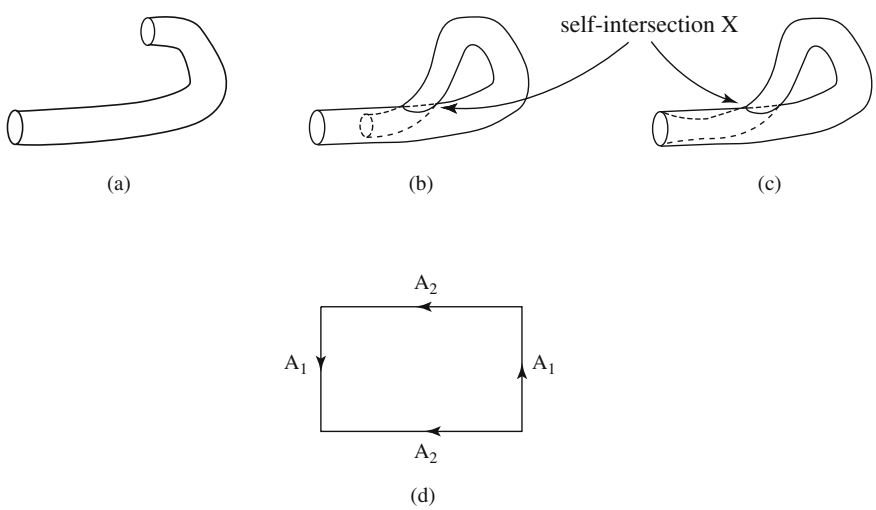



**Figure 6.24.** The Klein bottle.

gluing the ends together like we do for the torus, we bend one end and bring it to the inside of the tube (Figure 6.24(b)), and finally glue that end to the other one from the inside (Figure 6.24(c)). Of course, as can be seen from Figure 6.24, this involves piercing the tube. As a result we do not get a legitimate surface since we have a self-intersection along a circle (marked **X** in Figures 6.24(b) and (c)). Unfortunately, we cannot eliminate this singularity because the surface **K** that we were trying to define, like the projective plane, cannot be imbedded in $\mathbf{R}^3$. We are seeing **K** after it has been "pushed" into $\mathbf{R}^3$. On the other hand, if we had a fourth dimension, then we could grab one part of the surface near the intersection **X** and lift it up into the fourth dimension, thereby removing any intersection and creating a real surface without singularities. This is in analogy with the way that one can remove the self-intersection of a circle immersed in the plane as a figure eight. We can remove the self-intersection by lifting one part of the circle near the intersection up into the third dimension. See Figure 6.25. If one can understand how a two-dimensional person could try to visualize that three-dimensional construction, then one should be able visualize what the surface **K** looks like. It is clear from the construction, that **K** can also be described as a square with sides identified as shown in Figure 6.24(d). This means that we can give the following precise definition of a Klein bottle:

**Definition.** Any surface with symbol $A_1A_2A_1A_2^{-1}$ is called a *Klein bottle*.

**Proposition 6.5.11.** The Klein bottle **K** is a nonorientable surface that is homeomorphic to $\mathbf{P}^2 \# \mathbf{P}^2$.

**Proof.** Represent the Klein bottle via the labeled polygon shown in Figure 6.26. If we cut **K** along the lines indicated by **c** and **d**, then we get two Moebius strips with the shaded region being one of them. This shows that **K** is gotten by taking two Moebius strips and gluing them together along their boundaries. This is precisely what we have in the case of $\mathbf{P}^2 \# \mathbf{P}^2$, since the projective plane can be gotten by gluing a disk to a Moebius strip along their boundaries.

Intuitively, we can see that the Klein bottle is nonorientable because it is one-sided. Here we are thinking of the surface as made of, say, paper and we are walking along it. See our discussion of the Moebius strip in Section 1.6. Because of the one-

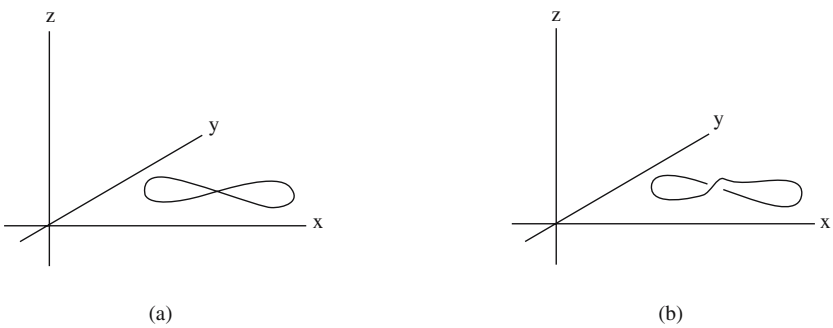



**Figure 6.25.** Eliminating a self-intersection using a third dimension.

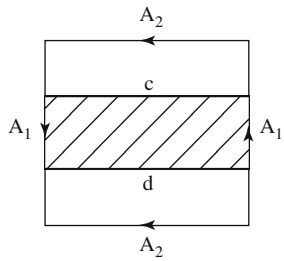



**Figure 6.26.** The Klein bottle as the union of two Moebius strips.

sidedness, one could say that if one were to pour water "into" a Klein bottle, then the water would come right "out," although this is not quite true since gravity would keep some of it inside.

Sometimes a Klein bottle is referred to as a sphere with a twisted handle and the connected sum of k Klein bottles, as a *sphere with k twisted handles*.

## 6.6 Bordered and Noncompact Surfaces

We finish the chapter with a few comments on "bordered" and noncompact surfaces. To allow for a boundary we broaden the definition of a combinatorial surface.

**Definition.** A *combinatorial surface* (with or without boundary) is a polyhedron **S** that admits a triangulation $(K,\varphi)$ satisfying:

(1) K is a two-dimensional connected simplicial complex.
(2) Each 1-simplex of K is a face of at least one but not more than two 2-simplices of K.
(3) For every vertex **v** in K, the distinct 2-simplices $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_s$ of K to which **v** belongs can be ordered in such a way that $\boldsymbol{\sigma}_i$, $1 \le i \le s-1$, meets $\boldsymbol{\sigma}_{i+1}$ in precisely one 1-simplex. If $s > 1$, then $\boldsymbol{\sigma}_s$ and $\boldsymbol{\sigma}_1$ have either just **v** or a single 1-simplex in common.

The *boundary* of **S**, denoted by $\partial\mathbf{S}$, is defined to be the set $\varphi(|\partial K|)$. If $\partial\mathbf{S} = \boldsymbol{\phi}$, then **S** is called a *closed* surface. A surface with nonempty boundary is called a *surface with boundary* or a *bordered surface*.

Clearly, if **S** is a combinatorial surface using the new definition and if $\partial\mathbf{S} = \boldsymbol{\phi}$, then **S** is just a combinatorial surface as defined earlier in Section 6.5. As before, compactness and connectedness is built into the definition of these new surfaces. We shall again drop the adjective "combinatorial."

Topologically a bordered surface can be represented by a polygon some, but **not** necessarily all, of whose sides have been identified in pairs. The unidentified sides of the polygon give rise to the boundary of the surface. Therefore, to classify bordered surfaces one can take the same approach as the one used to classify closed surfaces. One shows that every bordered surface can be represented by a labeled polygon and that such representatives can be put into a "normal form."

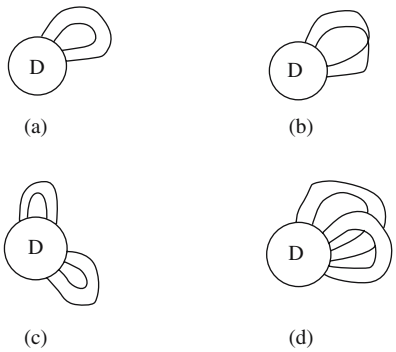



**Figure 6.27.** Bordered surfaces as a disk **D** with handles.

**6.6.1. Theorem.** (The Classification Theorem for Bordered Surfaces) Two bordered surfaces are homeomorphic if and only if they have the same number of boundary components, they are both orientable or nonorientable, and they have the same Euler characteristic.

**Proof.** See [AgoM76].

Theorem 6.6.1 tells us when two bordered surfaces are homeomorphic, but what does the typical one look like? To answer that question, we need an easy to understand list of bordered surfaces to which an arbitrary one can be compared. A simple construction produces such a list. Start with a disk and some rectangular strips. Successively paste the two ends of these strips to the boundary of the disk. The bordered surface that one gets in this way is called a "disk with handles." See Figure 6.27. Note that each strip can be attached in different ways. Before we paste the ends to the disk, we have the option of either giving or not giving the strip a half-twist (see Figure 6.27(a) and (b)). The ends themselves can be pasted in a manner so that the strip either does or does not interlock with previous strips (compare Figures 6.27(c) and (d)). By computing the number of boundary components, the orientability, and the Euler characteristic of such disks with handles, one can show, using Theorem 6.6.1, that every bordered surface is homeomorphic to one of these (see [Mass67]). One corollary of this that all bordered surfaces can be imbedded in $\mathbf{R}^3$. This was not true for closed surfaces.

Finally, the classification of **non**compact surfaces with or without boundary turns out to be much more difficult than that of compact surfaces because the number of possibilities is so much greater. Some examples of noncompact surfaces are:

(1) Any open subset of a compact surface (alternatively, the complement of an arbitrary closed set).
(2) The surface of a ladder with an infinite number of rungs.
(3) The surface of an infinite wire grid that is infinite in both directions.
(4) Infinite connected sums of surfaces.

We refer the interested reader to [Mass67]. The classification results are much messier than those in the compact case. Note that all the triangulations are infinite in this case.

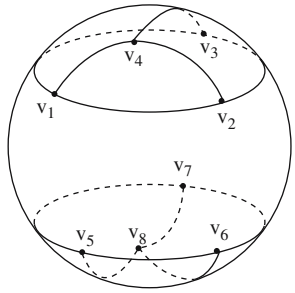



**Figure 6.28.** A decomposition of $\mathbf{S}^2$ for which Euler's theorem fails.

## 6.7 Exercises

### Section 6.1

6.1.1. The proof of Euler's formula (Theorem 6.1.1) used the fact that the triangular faces in the decomposition of a disk $\mathbf{D}^2$ can be listed in a sequence $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_k$, such that $\mathbf{T}_i$ meets

$$\mathbf{X}_{i-1} = \bigcup_{1 \le j < i} \mathbf{T}_j$$

in either one or two edges. Sketch a proof and discuss potential problems. (In [BurM71] it is proved that the analogous fact for cell decompositions of higher-dimensional disks does not hold.)

6.1.2. A common way to express Euler's theorem is to say that no matter how a sphere is divided into $n_f$ regions with $n_e$ edges and $n_v$ vertices, the sum $n_v - n_e + n_f$ will always equal 2. Compute $n_v - n_e + n_f$ for the decomposition of $\mathbf{S}^2$ shown in Figure 6.28. Where does the proof of Euler's theorem fail in that example? If we want to preserve the validity of the theorem, then what conditions must a "permissible region" satisfy so that Euler's theorem will hold for all decompositions of $\mathbf{S}^2$ into permissible regions?

### Section 6.3

6.3.1. Let K be a simplicial complex and let $\mathbf{x} \in |K|$. Prove that there is a unique simplex $\boldsymbol{\sigma} \in K$ such that $\mathbf{x} \in \operatorname{int} \boldsymbol{\sigma}$.

6.3.2. Show that if L and M are subcomplexes of a simplicial complex K, then $L \cap M$ is a subcomplex of K.

6.3.3. Prove that if K is a simplicial complex, then so is $\partial K$.

6.3.4. Let K be a simplicial complex.

(a) Prove that K is connected if and only if $|K|$ is connected.
(b) Define a *component* of K to be a maximal connected subcomplex L of K. Show that

$$K = L_1 \cup L_2 \cup \ldots \cup L_n,$$

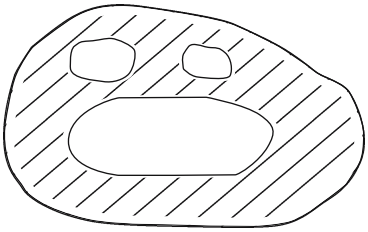


**Figure 6.29.** A bordered surface for Exercise 6.6.1.

where the $L_i$ are components of K. Show that $L_i \cap L_j = \phi$, if $i \neq j$.

(c) Prove that a subcomplex L of K is a component of K if and only if $|L|$ is a component of $|K|$.

6.3.5. Let K be a simplicial complex and let $f : |K| \to \mathbf{R}^m$ be a map. Prove that f is continuous if and only if $f|\sigma$ is continuous for all $\sigma \in K$.

6.3.6. Prove Proposition 6.3.3. (Hint: Use Exercise 6.3.5.)

## Section 6.5

6.5.1. Sketch a proof of the fact that a closed nonorientable surface cannot be imbedded in $\mathbf{R}^3$ by justifying and proving the following:

(a) A closed curve that meets a closed surface in $\mathbf{R}^3$ "transversally" must meet it in an even number of points.

(b) If a closed surface in $\mathbf{R}^3$ contains a Moebius strip, then there is a curve close to the median curve of the Moebius strip that meets the surface in only one point.

6.5.2. What surface has symbol $abcda^{-1}b^{-1}c^{-1}d^{-1}$?

6.5.3. Which of the surfaces listed in Proposition 6.5.6 is homeomorphic to the following:

(a) a connected sum of a torus and Klein bottle

(b) a connected sum of a torus and projective plane

Justify your answer.

6.5.4. The genus of a surface can be defined as the maximum number of disjoint circles along which a surface can be cut without disconnecting it. Justify this characterization of the genus.

6.5.5. If one is going to make computations with a triangulation of a space there are obvious reasons for choosing a *minimal* triangulation, that is, a triangulation that has the fewest number of simplices. Let **S** be a surface triangulated by a simplicial complex K. Prove the following lower bounds on $n_i = n_i(K)$, the number of i-simplices in K:

(a) $\mathbf{S} = \mathbf{S}^2$ : $n_0 \geq 4$, $n_1 \geq 6$, $n_2 \geq 4$

(b) $\mathbf{S} = \mathbf{P}^2$ : $n_0 \geq 6$
$n_1 \geq 15$
$n_2 \geq 10$

(c) $\mathbf{S} = \mathbf{S}^1 \times \mathbf{S}^1$ : $n_0 \geq 7$
$n_1 \geq 21$
$n_2 \geq 14$

**Hint:** Use facts about the Euler characteristic and relations between the numbers $n_i$ as was done in the proof of Theorem 6.1.2.

## Section 6.6

6.6.1. Triangulate the bordered surface shown in Figure 6.29. What is its Euler characteristic? Represent the surface as a disk with handles.

CHAPTER 7

# Algebraic Topology

## 7.1 Introduction

The central problem of algebraic topology is to classify spaces up to homeomorphism by means of computable algebraic invariants. In the last chapter we showed how two invariants, namely, the Euler characteristic and orientability, gave a complete classification of surfaces. Unfortunately, these invariants are quite inadequate to classify higher-dimensional spaces. However, they are simple examples of the much more general invariants that we shall discuss in this chapter.

The heart of this chapter is its introduction to homology theory. Section 7.2.1 defines the homology groups for simplicial complexes and polyhedra, and Section 7.2.2 shows how continuous maps induce homomorphisms of these groups. Section 7.2.3 describes a few immediate applications. In Section 7.2.4 we indicate how homology theory can be extended to cell complexes and how this can greatly simplify some computations dealing with homology groups. Along the way we define CW complexes, which are really the spaces of choice in algebraic topology because one can get the most convenient description of a space with them. Section 7.2.5 defines the incidence matrices for simplicial complexes. These are a fundamental tool for computing homology groups with a computer. Section 7.2.6 describes a useful extension of homology groups where one uses an arbitrary coefficient group, in particular, $\mathbf{Z}_2$. After this overview of homology theory we move on to define cohomology in Section 7.3. The cohomology groups are a kind of dual to the homology groups. We then come to the other major classical topic in algebraic topology, namely, homotopy theory. We start in Sections 7.4.1 and 7.4.2 with a discussion of the fundamental group of a topological space and covering spaces. These topics have their roots in complex analysis. Section 7.4.3 sketches the definition of the higher-dimensional homotopy groups and concludes with some major theorems from homotopy theory. Section 7.5 is devoted to pseudomanifolds, the degree of a map, manifolds, and Poincaré duality (probably the single most important algebraic property of manifolds and the property that sets manifolds apart from other spaces). We wrap up our overview of algebraic topology in Section 7.6 by telling the reader briefly about important aspects that we did not have time for and indicate further topics to pursue. Finally, as one last example,

Section 7.7 applies the theory developed in this chapter to our ever-interesting space $\mathbf{P}^n$.

The reader is warned that this chapter may be especially hard going if he/she has not previously studied some abstract algebra. We shall not be using any really advanced ideas from abstract algebra, but if the reader is new to it and has no one for a guide, then, as usual, it will take a certain amount of time to get accustomed to thinking along these lines. Groups and homomorphism are quite a bit different from topics in calculus and basic linear algebra. The author hopes the reader will persevere because in the end one will be rewarded with some beautiful theories. The next chapter will make essential use of what is developed here and apply it to the study of manifolds. Manifolds are the natural spaces for geometric modeling and getting an understanding of our universe.

Anyone reading this chapter should at least read Sections 7.2.1–7.2.5 as carefully as possible. They give the reader uninitiated to algebraic topology a taste of what the subject is about. It is important to pay attention to the details and work through them, otherwise little will sink in and everything will be just a blur. Two other important topics are the fundamental group and pseudomanifolds. Although we try to be as clear as we can be about basic concepts, the proofs in the chapter will tend to get less and less detailed as we go along because we want to give as much of an overview as possible. Even if things start getting too abstract, it is recommended that one glance over all the material (definitions and basic theorems) to at least get an overall picture of how algebraic topology tries to pin down the structure and classification of topological spaces.

## 7.2 Homology Theory

### 7.2.1 Homology Groups

The motivation for the homology groups (and especially the homotopy groups defined later in the chapter) is based on the intuitive idea that topological spaces can be characterized in terms of the number and type of "holes" that they have. There is no precise definition of a hole. The simplest examples of spaces without holes are the Euclidean spaces $\mathbf{R}^n$. The prototype of a space that has an "n-dimensional hole" is $\mathbf{S}^n$. The hole exists because $\mathbf{S}^n$ cannot be contracted to a point **within** $\mathbf{S}^n$ without tearing it. In our discussion we shall use surfaces and their one-dimensional holes as examples because one can draw nice simple pictures in this case to illustrate what we are talking about.

Consider the infinite cylinder $\mathbf{C}$ in Figure 7.1. In terms of holes, we would say that the cylinder has a "vertical hole." The existence of this hole is demonstrated by the fact that we have closed curves (circles), such as $\alpha$ and $\beta$, that cannot be contracted to a point **in** $\mathbf{C}$. The curves $\alpha$ and $\beta$ actually determine the same hole in $\mathbf{C}$. One reason for this is that one can deform one curve into the other. On the other hand, the closed curves $\alpha$ and $\beta$ in the torus $\mathbf{T}$ in Figure 7.1 correspond to the presence of two distinct holes, an "inside" and "outside" hole. Of course, some closed curves, such as $\gamma$ in the cylinder $\mathbf{C}$, can be contracted to a point and do not correspond to any hole. It follows that if we are going to study (one-dimensional) holes by means of closed curves then

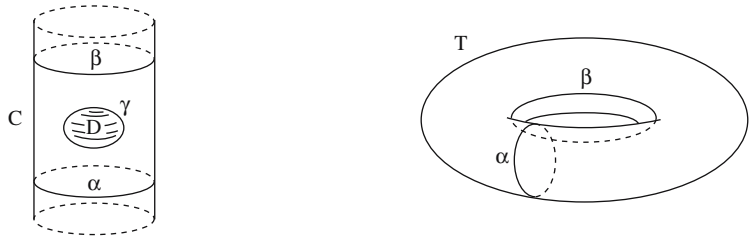



**Figure 7.1.** Holes in a torus.

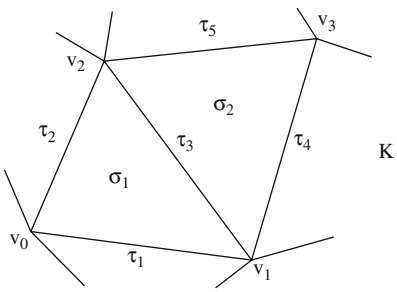



**Figure 7.2.** Boundaries of simplices in complexes.

we need to look at some sort of equivalence classes of closed curves with respect to a suitable equivalence relation. One natural such equivalence relation would be homotopy. This is what one uses for the definition of homotopy groups. For homology groups we shall use a weaker and more algebraic notion.

Note that the curve $\boldsymbol{\gamma}$ in $\mathbf{C}$ bounds a disk $\mathbf{D}$ and that the union of the two curves $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in $\mathbf{C}$ also bounds a region. This suggests that we define an equivalence relation for closed curves where the equivalence class that represents the trivial element corresponds to curves that bound. We continue this line of thought in the context of simplicial complexes. One of the useful aspects of simplicial complexes is that they have both a geometric and an abstract nature to them. It is the latter that we want to take advantage of right now because it will let us switch into a symbol manipulation mode where purely algebraic manipulations replace geometric operations.

Figure 7.2 shows two 2-simplices, $\boldsymbol{\sigma}_1 = \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2$ and $\boldsymbol{\sigma}_2 = \mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$, in a simplicial complex K. Consider the edges $\boldsymbol{\tau}_1 = \mathbf{v}_0\mathbf{v}_1$, $\boldsymbol{\tau}_2 = \mathbf{v}_0\mathbf{v}_2$, $\boldsymbol{\tau}_3 = \mathbf{v}_1\mathbf{v}_2$, $\boldsymbol{\tau}_4 = \mathbf{v}_1\mathbf{v}_3$, and $\boldsymbol{\tau}_5 = \mathbf{v}_2\mathbf{v}_3$ in K. As point sets, the boundaries of $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_1 \cup \boldsymbol{\sigma}_2$ are $\boldsymbol{\tau}_1 \cup \boldsymbol{\tau}_2 \cup \boldsymbol{\tau}_3$, $\boldsymbol{\tau}_3 \cup \boldsymbol{\tau}_4 \cup \boldsymbol{\tau}_5$, and $\boldsymbol{\tau}_1 \cup \boldsymbol{\tau}_2 \cup \boldsymbol{\tau}_4 \cup \boldsymbol{\tau}_5$, respectively. What is the relationship between the boundary of the region $\boldsymbol{\sigma}_1 \cup \boldsymbol{\sigma}_2$ and the boundaries of the individual simplices $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$? We could say that the edge $\boldsymbol{\tau}_3$, considered as part of the boundary of $\boldsymbol{\sigma}_1$, and the edge $\boldsymbol{\tau}_3$, considered as part of the boundary of $\boldsymbol{\sigma}_2$, have, in some sense, cancelled each other. This idea of cancellation can be made more meaningful if we use a more suggestive algebraic notation and write "+" instead of "$\cup$." If we want the function "boundary of" to be additive, then we want the equation

$$\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + \boldsymbol{\tau}_4 + \boldsymbol{\tau}_5 = (\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + \boldsymbol{\tau}_3) + (\boldsymbol{\tau}_3 + \boldsymbol{\tau}_4 + \boldsymbol{\tau}_5)$$

to hold. If we treat simplices as formal symbols, then one way to satisfy this equation is to identify the expression $\boldsymbol{\tau}_3 + \boldsymbol{\tau}_3$ (which one is also tempted to write as $2\boldsymbol{\tau}_3$) with 0. A homology theory based on this approach (the mod 2 homology groups) will be discussed in Section 7.2.6. In this section we describe a second approach. It may seem slightly more complicated initially but will lead to better invariants.

First of all, we introduce a notion of orientation for simplices. For now, we shall leave things to intuition and infer the orientation of a simplex simply from the order in which its vertices are listed. For example, the expression $\mathbf{vw}$ will be used to determine not only the 1-simplex with vertices $\mathbf{v}$ and $\mathbf{w}$ but also the direction (in this case from $\mathbf{v}$ to $\mathbf{w}$) in which one should "walk" if one were to walk along that edge. Given an orientation of a 2-simplex, we get a natural orientation of the boundary curve. For example, returning to the simplicial complex in Figure 7.2, using the expression $\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2$ for the 2-simplex $\boldsymbol{\sigma}_1$ indicates that it has been oriented in the counterclockwise fashion and its boundary will be thought of as a closed path that is intended to be traversed also in a counterclockwise fashion, say by starting at $\mathbf{v}_0$ and then walking from $\mathbf{v}_0$ to $\mathbf{v}_1$, from $\mathbf{v}_1$ to $\mathbf{v}_2$, and finally from $\mathbf{v}_2$ back to $\mathbf{v}_0$.

We can express the relationships between oriented simplices and their boundaries symbolically by introducing a boundary operator $\partial^*$, so that what we were just saying can be summarized by equations of the form

$$\partial^*(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2) = \mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_0$$

and

$$\partial^*(\mathbf{v}_1\mathbf{v}_3\mathbf{v}_2) = \mathbf{v}_1\mathbf{v}_3 + \mathbf{v}_3\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_1.$$

More generally, since arbitrary oriented regions also define an orientation on their boundary, it makes sense to have the operator $\partial^*$ defined on those. In the case of the union of the two simplices $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ the geometry would imply that

$$\partial^*(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 \cup \mathbf{v}_1\mathbf{v}_3\mathbf{v}_2) = \mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_1\mathbf{v}_3 + \mathbf{v}_3\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_0. \tag{7.1}$$

On the other hand, from an algebraic point of view we would like

$$\begin{aligned}\partial^*(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_1\mathbf{v}_3\mathbf{v}_2) &= \partial^*(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2) + \partial^*(\mathbf{v}_1\mathbf{v}_3\mathbf{v}_2)\\ &= (\mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_0) + (\mathbf{v}_1\mathbf{v}_3 + \mathbf{v}_3\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_1).\end{aligned} \tag{7.2}$$

The difference between the two expressions on the right of equations (7.1) and (7.2) is that the second has an extra $\mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_1$ term. In terms of walking along paths, the difference between the two paths is that in the second we took extra strolls, first from $\mathbf{v}_1$ to $\mathbf{v}_2$ and later from $\mathbf{v}_2$ back to $\mathbf{v}_1$. This suggests that we should identify $\mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_1$ with 0. By formally defining $\mathbf{v}_2\mathbf{v}_1 = -\mathbf{v}_1\mathbf{v}_2$ we make this even more plausible. Geometrically, it means that $-\mathbf{v}_1\mathbf{v}_2$ represents the path from $\mathbf{v}_1$ to $\mathbf{v}_2$ traversed in the opposite direction. We would then have the reasonable looking equations

$$\mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_1 = \mathbf{v}_1\mathbf{v}_2 + (-\mathbf{v}_1\mathbf{v}_2) = 0.$$

This as far as we go in our motivational discussion and we now start our rigorous development of homology groups. What we tried to indicate was that if one is interested in studying the holes in a space, then one approach to this leads to symbol manipulation involving oriented simplices, formal linear sums of these, and boundary maps. A homology group will be a "group of cycles" modulo a "group of boundaries."

Let $\boldsymbol{\sigma}$ be a q-simplex in $\mathbf{R}^n$. An orientation of $\boldsymbol{\sigma}$ cannot just be an ordering of the set V of its vertices because there are many orderings. We need an equivalence relation on the set of orderings. Note that any two orderings of V differ by a permutation of V.

**Definition.** An *orientation* of $\boldsymbol{\sigma}$ is an equivalence class of orderings of the vertices of $\boldsymbol{\sigma}$, where two orderings are said to be equivalent if they differ by an even permutation of the vertices.

The fact that the inverse of an even permutation is an even permutation easily implies that we are dealing with an equivalence relation. Furthermore, there are only two equivalence classes. The two possible orientations of a q-simplex $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q$, $q \geq 1$ are determined by the orderings $(\mathbf{v}_0,\mathbf{v}_1, \ldots ,\mathbf{v}_q)$ and $(\mathbf{v}_1,\mathbf{v}_0,\mathbf{v}_2,\mathbf{v}_3, \ldots ,\mathbf{v}_q)$. If $\mu$ is one orientation of $\boldsymbol{\sigma}$, then it will be convenient to let $-\mu$ denote the other. we have $-(-\mu) = \mu$. A 0-simplex has only one orientation. Note the similarity between the definition of the orientation of a simplex with the definition of the orientation of a vector space.

**Definition.** An *oriented q-simplex* $[\boldsymbol{\sigma}]$ is a pair $(\boldsymbol{\sigma},\mu)$, where $\boldsymbol{\sigma}$ is a q-simplex and $\mu$ is an orientation of $\boldsymbol{\sigma}$. The notation $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$ denotes the oriented q-simplex $(\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q,\mu)$ where $\mu$ the orientation determined by the ordering $(\mathbf{v}_0,\mathbf{v}_1, \ldots , \mathbf{v}_q)$. If $q = 0$, then there is only one orientation, and we shall always write simply $\mathbf{v}_0$ instead of $[\mathbf{v}_0]$. If $q \geq 1$ and if $[\boldsymbol{\sigma}]$ is an oriented q-simplex, then $-[\boldsymbol{\sigma}]$ is defined to be the oriented q-simplex consisting of $\boldsymbol{\sigma}$ together with the opposite orientation, that is, if $[\boldsymbol{\sigma}] = [\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$, then $-[\boldsymbol{\sigma}] = [\mathbf{v}_1\mathbf{v}_0\mathbf{v}_2\mathbf{v}_3 \cdots \mathbf{v}_q]$. For uniformity of notation, $[\boldsymbol{\sigma}]$ may also be denoted by $+[\boldsymbol{\sigma}]$.

Now, let K be a simplicial complex and let $S_q$ denote the set of oriented q-simplices of K.

**Definition.** A *q-chain* of K, $0 \leq q \leq \dim K$, is a function $f: S_q \to \mathbf{Z}$ with the additional property that if $q \geq 1$, then

$$f(-[\boldsymbol{\sigma}^q]) = -f([\boldsymbol{\sigma}^q])$$

for every $[\boldsymbol{\sigma}^q]$ in $S_q$. The set of all q-chains of K is denoted by $C_q(K)$. Given $f, g \in C_q(K)$, define the sum

$$f + g: S_q \to \mathbf{Z}$$

by

$$(f + g)([\boldsymbol{\sigma}^q]) = f([\boldsymbol{\sigma}^q]) + g([\boldsymbol{\sigma}^q]).$$

**7.2.1.1. Theorem.** $(C_q(K),+)$ is an abelian group.

**Proof.** First of all, it is easy to see that $f + g \in C_q(K)$. The additive identity for + is the zero function, which maps all oriented q-simplices to zero. The additive inverse of any f in $C_q(K)$, denoted by $-f$, is defined by the formula

$$(-f)([\sigma^q]) = -(f([\sigma^q]))$$

for all $[\sigma^q] \in S_q$.

It is convenient to have $C_q(K)$ defined for all values of q, including negative values, even though the only groups that are really interesting to us are the ones with $0 \le q \le \dim K$.

**Definition.** If $q < 0$ or $q > \dim K$, define $C_q(K) = 0$.

**Definition.** For all q, the abelian group $(C_q(K),+)$ is called the *group of q-chains* of K. The "vector"

$$C_\#(K) = (\ldots, C_{-1}(K), C_0(K), C_1(K), \ldots)$$

is called the (*oriented*) *chain complex* of K.

The definition of q-chains as functions is neither convenient nor intuitive. We shall now describe the more common notation that one uses when working with q-chains. For each oriented q-simplex $\alpha = [\sigma^q] \in S_q$ define a q-chain

$$\alpha_F \in C_q(K)$$

by

$$\begin{aligned} \alpha_F(\beta) &= 0, \quad \text{for} \quad \beta \in S_q \quad \text{and} \quad \beta \ne \pm\alpha, \\ \alpha_F(\alpha) &= 1, \quad \text{and} \\ \alpha_F(-\alpha) &= -1. \end{aligned}$$

Such "elementary" q-chains $\alpha_F$ actually generate $C_q(K)$. To see this, choose one orientation for each q-simplex of K and let $S_q^+$ be the collection of oriented q-simplices defined by these choices. (The easiest way to simultaneously pick an orientation for all the simplices of K is to order the vertices of K once and for all and then to take the induced orientation.) The essential property of $S_q^+$ is that it is a subset of $S_q$ satisfying:

(1) If $q = 0$, then $S_q^+ = S_q$.
(2) If $q \ge 1$, then, for any $\beta \in S_q$, either $\beta$ or $-\beta$ belongs to $S_q^+$ but **not** both.

**7.2.1.2. Lemma.** $C_q(K) = \bigoplus_{\alpha \in S_q^+} \mathbf{Z}\alpha_F$.

**Proof.** This is easy to prove (Exercise 7.2.1.1). See Appendix B for a clarification of the notation. A proof can also be found in [AgoM76].

Because the map

$$\begin{aligned} \mathbf{Z} &\to \mathbf{Z}\alpha_F \\ n &\to n\alpha_F \end{aligned}$$

is clearly an isomorphism, it follows from Lemma 7.2.1.2 that if K has $n_q$ q-simplices, then $C_q(K)$ is isomorphic to a free abelian group, which is a direct sum of $n_q$ copies

of **Z**. Since the map $\alpha \to \alpha_F$ imbeds $S_q$ in $C_q(K)$, we shall identify $\alpha$ with $\alpha_F$. With this identification, we now have a rigorous mathematical definition of the notion, referred to in the motivational part at the beginning of this section, of "formal linear combinations" of oriented q-simplices. Furthermore, by treating the elements of $C_q(K)$ as such formal sums, which we shall do in the future, we shall make our computations more intuitive. The reader should remember, however, that the definition of $C_q(K)$ depends only on K and not on any particular choice of orientations.

Now that we know about q-chains, we move on to a definition of the boundary map.

**Definition.** The *boundary map*

$$\partial_q : C_q(K) \to C_{q-1}(K)$$

is defined as follows:

(1) If $1 \le q \le \dim K$, then $\partial_q$ is the unique homomorphism with the property that

$$\partial_q([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) = \sum_{i=0}^{q} (-1)^i [\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]$$

for each oriented q-simplex $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$ of K, where "$\hat{\mathbf{v}}_i$" denotes the fact that the vertex $\mathbf{v}_i$ has been omitted.

(2) If $q \le 0$ or $q > \dim K$, then $\partial_q$ is defined to be the zero homomorphism.

**7.2.1.3. Lemma.**

(1) The maps $\partial_q$ are well-defined homomorphisms.
(2) For all q, $\partial_{q-1} \circ \partial_q = 0$.

**Proof.** Assume that $1 \le q \le \dim K$, which is the only case where something has to be proved. Let $\boldsymbol{\sigma}$ be a q-simplex and $\boldsymbol{\sigma}_i$ a $(q-1)$-dimensional face of $\boldsymbol{\sigma}$. Suppose that

$$\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q \quad \text{and} \quad \boldsymbol{\sigma}_i = \mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q.$$

Let $[o]$ and $[o_i]$ be the orientations of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_i$ induced by the orderings

$$o = (\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_q) \quad \text{and} \quad o_i = (\mathbf{v}_0, \ldots, \hat{\mathbf{v}}_i, \ldots, \mathbf{v}_q),$$

respectively.

**Claim.** The orientation $\nu = (-1)^i[o_i]$ of $\boldsymbol{\sigma}_i$ depends only on the orientation $\mu = [o]$ and not on the particular ordering o.

First, consider what happens to the orientation $[o_i]$ when we pass from the ordering o to the ordering

$$o' = (\mathbf{v}_0, \ldots, \mathbf{v}_t, \ldots, \mathbf{v}_s, \ldots, \mathbf{v}_q), \quad s < t,$$

which corresponds to interchanging two vertices $\mathbf{v}_s$ and $\mathbf{v}_t$. If $s \ne i \ne t$, then we have interchanged two vertices of $\boldsymbol{\sigma}_i$, so that $[o'_i] = -[o_i]$ and

$$\nu' = (-1)^i\left[o_i'\right] = -(-1)^i[o_i] = -\nu.$$

If s = i, then the new ordering of the vertices of $\boldsymbol{\sigma}$ is

$$(\mathbf{v}_0, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_t, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{t-1}, \mathbf{v}_{t+1} \ldots, \mathbf{v}_q)$$

and

$$\nu' = (-1)^t\left[o_i'\right] = (-1)^t(-1)^{t-i-1}[o_i] = -(-1)^i[o_i] = -\nu.$$

A similar equation holds if t = i. Therefore, interchanging two distinct vertices of o always results in a change of sign of $\nu$. Since the representative o for $\mu$ is well defined up to an even permutation of the vertices and since every even permutation is the composition of an even number of transpositions, the Claim is proved.

**Definition.** The orientation $\nu$ of $\boldsymbol{\sigma}_i$ is called the *orientation* of $\boldsymbol{\sigma}_i$ *induced by the orientation* $\mu$ *of* $\boldsymbol{\sigma}$.

The Claim implies that $\partial_q$ is well defined because the "boundary" (q – 1)-chain

$$\sum_{i=0}^{q}(-1)^i[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]$$

is well defined for each oriented q-simplex $[\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]$. Furthermore, because this (q – 1)-chain is exactly what the (oriented) boundary of the oriented q-simplex should be intuitively, we are justified in calling the maps $\partial_q$ "boundary maps." For example, the definition of $\partial_2$ implies that

$$\begin{aligned}\partial_2([\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2]) &= [\mathbf{v}_1\mathbf{v}_2] - [\mathbf{v}_0\mathbf{v}_2] + [\mathbf{v}_0\mathbf{v}_1]\\ &= [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_0] + [\mathbf{v}_0\mathbf{v}_1],\end{aligned}$$

which is what we want. Part (1) of the lemma is now proved since to define a homomorphism on a free group it suffices to specify it on a basis (Theorem B.5.9).

To prove part (2), we may assume that $q \geq 2$. It suffices to show that

$$(\partial_{q-1}\circ\partial_q)([\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]) = 0$$

for every oriented q-simplex $[\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]$ of K, since these generate $C_q(K)$. Now

$$\begin{aligned}(\partial_{q-1}\circ\partial_q)([\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]) &= \partial_{q-1}\left(\sum_{i=0}^{q}(-1)^i[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\right)\\ &= \sum^{q}(-1)^i\partial_{q-1}([\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q])\\ &= \sum_{i=0}^{q}(-1)^i\Bigg(\sum_{j=0}^{i-1}(-1)^j[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_j \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\\ &\qquad + \sum_{j=i+1}^{q}(-1)^{j-1}[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \hat{\mathbf{v}}_j \cdots \mathbf{v}_q]\Bigg).\end{aligned}$$

Observe that, for $s < t$, each term $[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_s \cdots \hat{\mathbf{v}}_t \cdots \mathbf{v}_q]$ appears twice in the sum above. It appears once with coefficient $(-1)^{i+j-1}$ when $i = s$ and $j = t$, and a second time with coefficient $(-1)^{i+j}$ when $j = s$ and $i = t$. It follows that the terms in the sum cancel pairwise, so that the sum is zero. The lemma is proved.

We shall see that the second part of Lemma 7.2.1.3 is fundamental to the whole theory of homology groups. Here are some more basic definitions.

**Definition.** Let $c \in C_q(K)$. If $\partial_q(c) = 0$, then we shall call c a *q-cycle* of K. If $c = \partial_{q+1}(d)$ for some $d \in C_{q+1}(K)$, then we shall call c a *q-boundary* of K. The set of q-cycles and q-boundaries of K will be denoted by $Z_q(K)$ and $B_q(K)$, respectively.

Clearly,

$$Z_q(K) = \ker \partial_q \quad \text{and} \quad B_q(K) = \operatorname{im} \partial_{q+1},$$

so that we are dealing with subgroups of $C_q(K)$. Furthermore, by Lemma 7.2.1.3(2), the group $B_q(K)$ is actually a subgroup of $Z_q(K)$. It follows that we have inclusions

$$B_q(K) \subset Z_q(K) \subset C_q(K)$$

and it makes sense to talk about the quotient group of q-cycles modulo q-boundaries.

**Definition.** The *q-th homology group* of K, $H_q(K)$, is defined by

$$H_q(K) = \frac{Z_q(K)}{B_q(K)}.$$

As in the case of $C_q(K)$, it is notationally convenient to have the groups $B_q(K)$, $Z_q(K)$, and $H_q(K)$ defined for all values of q. Of course, only values of q satisfying $0 \le q \le \dim K$ are interesting. Negative values of q will in particular always be ignored in computations.

With the definition of the homology groups we have arrived at some important algebraic invariants for polyhedra, although we shall have to establish quite a few other facts before we will be ready to prove this. We could have used other regular figures, such as q-dimensional cubes, as building blocks for spaces to define these groups but there are at least two reasons for the choice of simplices, namely, with other figures both the orientation and the important maps $\partial_q$ would have been more complicated to define.

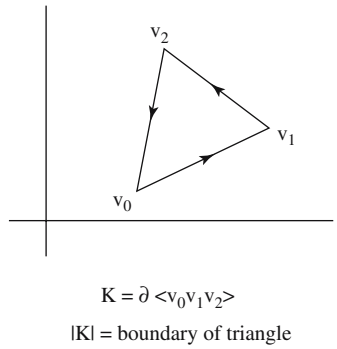
**7.2.1.4. Example.** To compute the homology groups of $K = \{\mathbf{v}_0\}$.

**Solution.** Clearly,

$$\begin{aligned} C_q(K) &= 0, && \text{if } q > 0, \\ C_0(K) &= \mathbf{Z}\mathbf{v}_0, && \text{and} \\ B_0(K) &= 0, \end{aligned}$$

so that

**Figure 7.3.** A triangulating complex for a circle.




$$H_q(K) = Z_q(K) = B_q(K) = 0, \quad \text{if } q > 0, \quad \text{and}$$
$$H_0(K) = Z_0(K) = C_0(K) \approx \mathbf{Z}.$$

**7.2.1.5. Example.** To compute the homology groups of $K = \partial < \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 >$. See Figure 7.3.

**Solution.** In this case

$$\begin{aligned} C_q(K) &= 0, \qquad \text{if } q > 1, \\ C_1(K) &= \mathbf{Z}[\mathbf{v}_0\mathbf{v}_1] \oplus \mathbf{Z}[\mathbf{v}_1\mathbf{v}_2] \oplus \mathbf{Z}[\mathbf{v}_2\mathbf{v}_0], \quad \text{and} \\ C_0(K) &= \mathbf{Z}\mathbf{v}_0 \oplus \mathbf{Z}\mathbf{v}_1 \oplus \mathbf{Z}\mathbf{v}_2. \end{aligned}$$

Trivially,

$$H_q(K) = Z_q(K) = B_q(K) = 0, \quad \text{if } q > 1.$$

Next, assume that

$$x = a[\mathbf{v}_0\mathbf{v}_1] + b[\mathbf{v}_1\mathbf{v}_2] + c[\mathbf{v}_2\mathbf{v}_0]$$

is a 1-cycle. Then

$$\begin{aligned} 0 &= \partial_1(x) \\ &= (a\mathbf{v}_1 - a\mathbf{v}_0) + (b\mathbf{v}_2 - b\mathbf{v}_1) + (c\mathbf{v}_0 - c\mathbf{v}_2) \\ &= (c - a)\mathbf{v}_0 + (a - b)\mathbf{v}_1 + (b - c)\mathbf{v}_2. \end{aligned}$$

It follows that $c - a = a - b = b - c = 0$, which implies that $a = b = c$. Hence,

$$Z_1(K) = \mathbf{Z}([\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_0]) \approx \mathbf{Z}.$$

But $B_1(K) = 0$, and so

$$H_1(K) = Z_1(K) \approx \mathbf{Z}.$$

Before computing $H_0(K)$ it is convenient to introduce some more notation.

**Definition.** Let K be a simplicial complex. If x, y $\in C_q(K)$, then we shall say that *x is homologous to y*, or that *x and y are homologous*, and write x ~ y, provided that $x - y = \partial_{q+1}(w)$ for some $w \in C_{q+1}(K)$. Also, if $z \in Z_q(K)$, let $[z] = z + B_q(K) \in H_q(K)$. The coset [z] is called the *homology class* in $H_q(K)$ *determined by z*.

**Note.** Recall that if **v** is a vertex of K, then we use **v** to denote the element of $C_0(K)$ rather than [**v**]. Therefore, [**v**] will always mean the homology class of **v**.

The relation ~ on $C_q(K)$ is clearly an equivalence relation and two q-cycles x, y $\in Z_q(K) \subset C_q(K)$ determine the same homology class in $H_q(K)$ if and only if x ~ y.

Returning to the computation of $H_0(K)$ in Example 7.2.1.5, note that $Z_0(K) = C_0(K)$. Also,

$$\mathbf{v}_0 \sim \mathbf{v}_1 \sim \mathbf{v}_2$$

because

$$\partial_1([\mathbf{v}_0\mathbf{v}_1]) = \mathbf{v}_1 - \mathbf{v}_0 \quad \text{and} \quad \partial_1([\mathbf{v}_1\mathbf{v}_2]) = \mathbf{v}_2 - \mathbf{v}_1.$$

This proves that $H_0(K)$ is generated by the homology class $[\mathbf{v}_0]$. On the other hand, if $n[\mathbf{v}_0] = 0$, then

$$n\mathbf{v}_0 = \partial_1(x)$$

for some 1-chain

$$x = a[\mathbf{v}_0\mathbf{v}_1] + b[\mathbf{v}_1\mathbf{v}_2] + c[\mathbf{v}_2\mathbf{v}_0].$$

Applying $\partial_1$ to the element x and equating the coefficients of the vertices $\mathbf{v}_0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ that one gets to the same coefficients in $n\mathbf{v}_0 = n\mathbf{v}_0 + 0\mathbf{v}_1 + 0\mathbf{v}_2$ easily shows that n = 0. It follows that the map from **Z** to $H_0(K)$ that sends k to $k[\mathbf{v}_0]$ is an isomorphism, that is,

$$H_0(K) \approx \mathbf{Z}.$$

This finishes Example 7.2.1.5.

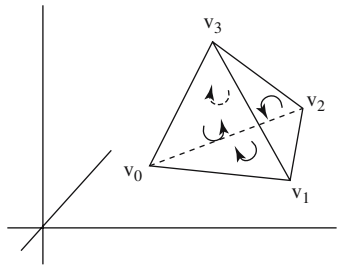
**7.2.1.6. Example.** To compute the homology groups of $K = \partial <\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3>$. See Figure 7.4.

**Solution.** First of all, since there are no q-simplices for q > 2,

$$H_q(K) = 0, \quad \text{if } q > 2.$$

On the other hand,

**Figure 7.4.** A triangulating complex for a sphere.




$K = \partial \langle v_0 v_1 v_2 v_3 \rangle$

$|K|$ = boundary of tetrahedron

$$C_2(K) = \mathbf{Z}[\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2] \oplus \mathbf{Z}[\mathbf{v}_0\mathbf{v}_1\mathbf{v}_3] \oplus \mathbf{Z}[\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3] \oplus \mathbf{Z}[\mathbf{v}_0\mathbf{v}_2\mathbf{v}_3],$$
$$C_1(K) = \mathbf{Z}[\mathbf{v}_0\mathbf{v}_1] \oplus \mathbf{Z}[\mathbf{v}_1\mathbf{v}_2] \oplus \mathbf{Z}[\mathbf{v}_0\mathbf{v}_2] \oplus \mathbf{Z}[\mathbf{v}_0\mathbf{v}_3] \oplus \mathbf{Z}[\mathbf{v}_3\mathbf{v}_2] \oplus \mathbf{Z}[\mathbf{v}_1\mathbf{v}_3], \quad \text{and}$$
$$C_0(K) = \mathbf{Z}\mathbf{v}_0 \oplus \mathbf{Z}\mathbf{v}_1 \oplus \mathbf{Z}\mathbf{v}_2 \oplus \mathbf{Z}\mathbf{v}_3.$$

Now $B_2(K) = 0$ and $\partial_0 = 0$. Therefore, $H_2(K) = Z_2(K)$ and $Z_0(K) = C_0(K)$. To compute the group $Z_2(K)$, let

$$z = a[\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2] + b[\mathbf{v}_0\mathbf{v}_1\mathbf{v}_3] + c[\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3] + d[\mathbf{v}_0\mathbf{v}_2\mathbf{v}_3]$$

be a 2-cycle. By definition, $\partial_2(z) = 0$. Computing the coefficients of the 1-simplices in $\partial_2(z)$ and setting them equal to 0 implies that $a = -b = -c = d$. In other words,

$$Z_2(K) = \mathbf{Z}([\mathbf{v}_0\mathbf{v}_2\mathbf{v}_1] \oplus [\mathbf{v}_0\mathbf{v}_1\mathbf{v}_3] \oplus [\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3] \oplus [\mathbf{v}_0\mathbf{v}_3\mathbf{v}_2]),$$

and so

$$H_2(K) \approx \mathbf{Z}.$$

We could compute $H_1(K)$ by calculating $Z_1(K)$ and $B_1(K)$ using arguments as before, but the approach to calculating homology groups by explicitly determining the group of cycles and the group of boundaries would become very tedious as spaces get more complicated. To simplify computations it is helpful to use certain tricks and shortcuts. The point is that we are looking for homology classes. Any representative cycle for such a class will do, and so we are free to replace any such cycle by a homologous one.

Let z be a 1-chain. The equation

$$[\mathbf{v}_2\mathbf{v}_3] = [\mathbf{v}_2\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_3] + \partial_2([\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3])$$

shows that the chains $[\mathbf{v}_2\mathbf{v}_3]$ and $[\mathbf{v}_2\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_3]$ are homologous. Therefore, if the oriented 1-simplex $[\mathbf{v}_2\mathbf{v}_3]$ appears in z with some nonzero coefficient, then we can replace $[\mathbf{v}_2\mathbf{v}_3]$ by $[\mathbf{v}_2\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_3]$ to get a homologous 1-chain $z_1$ in which $[\mathbf{v}_2\mathbf{v}_3]$ does not appear.

Similarly, we may replace any occurrence of $[\mathbf{v}_1\mathbf{v}_2]$ in $z_1$ by $[\mathbf{v}_1\mathbf{v}_0] + [\mathbf{v}_0\mathbf{v}_2]$. This will give us a 1-chain $z_2$, which is homologous to the original z and which contains neither $[\mathbf{v}_2\mathbf{v}_3]$ nor $[\mathbf{v}_1\mathbf{v}_2]$. If we assume further that z actually was a 1-cycle, then $z_2$ will be a 1-cycle and $[\mathbf{v}_0\mathbf{v}_2]$ cannot appear in $z_2$ either, otherwise the coefficient of $\mathbf{v}_2$ in $\partial_1(z_2)$ would not vanish. An argument similar to the one used in the computation of $H_1(K)$ in Example 7.2.1.5 now shows that $z_2$ must in fact be a multiple of

$$x = [\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_3] + [\mathbf{v}_3\mathbf{v}_0].$$

In other words, $H_1(K)$ is generated by [x]. On the other hand, $x = \partial_2([\mathbf{v}_0\mathbf{v}_1\mathbf{v}_3])$ and so $[x] = 0$ and we have proved that

$$H_1(K) = 0.$$

Finally, the group $H_0(K)$ can be computed like in Example 7.2.1.5. The first step is to show that $H_0(K)$ is generated by $[\mathbf{v}_0]$. Next, one shows that $n[\mathbf{v}_0] = 0$ implies that $n = 0$. We again get that

$$H_0(K) \approx \mathbf{Z}.$$

This finishes Example 7.2.1.6.

**7.2.1.7. Example.** The simplicial complex K in Figure 7.5 triangulates the torus. We want to compute the homology groups of K.

**Solution.** Again, since there are no q-simplices for $q > 2$,

$$H_q(K) = 0, \quad \text{if } q > 2.$$

To compute $H_2(K)$, orient the 2-simplices $\boldsymbol{\sigma}$ in K as indicated by the circular arrows in Figure 7.5. In this example, the notation $[\boldsymbol{\sigma}]$ will denote the 2-simplex $\boldsymbol{\sigma}$ together with that preferred orientation. Let $\Sigma$ denote the element in $C_2(K)$ that is the sum of these oriented 2-simplices, that is,

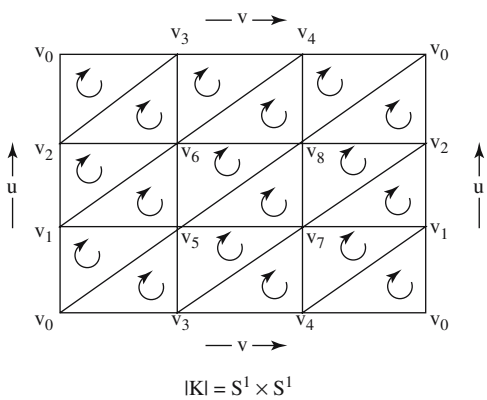



**Figure 7.5.** A triangulating complex for the torus.

$$\Sigma = \sum_{2\text{–simplex } \boldsymbol{\sigma} \text{ in } K} [\boldsymbol{\sigma}].$$

For example, $[\mathbf{v}_0\mathbf{v}_2\mathbf{v}_8]$ is one oriented simplex that appears in the sum $\Sigma$.

**Claim 1.** $\Sigma$ is a 2-cycle.

To see that $\partial_2(\Sigma) = 0$, consider an arbitrary 1-simplex $\mathbf{v}_i\mathbf{v}_j$ in K. If $\mathbf{v}_i\mathbf{v}_j$ is a face of the two 2-simplices $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, then the coefficients of $[\mathbf{v}_i\mathbf{v}_j]$ in $\partial_2([\boldsymbol{\sigma}])$ is the negative of the coefficient of $[\mathbf{v}_i\mathbf{v}_j]$ in $\partial_2([\boldsymbol{\sigma}'])$. It follows that the coefficient of $[\mathbf{v}_i\mathbf{v}_j]$ in $\partial_2(\Sigma)$ is zero.

**Claim 2.** $[\Sigma]$ generates $Z_2(K)$.

To prove Claim 2 let z be a 2-cycle. Let $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ be an arbitrary pair of 2-simplices in K that meet in an edge $\mathbf{v}_i\mathbf{v}_j$. It is easy to see that the coefficient of $[\mathbf{v}_i\mathbf{v}_j]$ will vanish in $\partial_2(z)$ if and only if $[\boldsymbol{\sigma}]$ and $[\boldsymbol{\sigma}']$ appear in z with the same multiplicity. Since this is true for all pairs of adjacent 2-simplices in K, we must have that $z = a\Sigma$ for some integer a.

Claim 1 and 2 prove that

$$H_2(K) = Z_2(K) = \mathbf{Z}\Sigma \approx \mathbf{Z}.$$

To determine $H_1(K)$, define two 1-cycles u and v by

$$u = [\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_0] \quad \text{and} \quad v = [\mathbf{v}_0\mathbf{v}_3] + [\mathbf{v}_3\mathbf{v}_4] + [\mathbf{v}_4\mathbf{v}_0].$$

**Claim 3.** $H_1(K)$ is generated by [u] and [v].

To prove Claim 3 observe that **any** 1-chain z is homologous to a 1-chain $z_1$ of the form

$$\begin{aligned} z_1 = {} & a_1[\mathbf{v}_0\mathbf{v}_3] + a_2[\mathbf{v}_3\mathbf{v}_4] + a_3[\mathbf{v}_4\mathbf{v}_0] + a_4[\mathbf{v}_0\mathbf{v}_1] + a_5[\mathbf{v}_1\mathbf{v}_2] \\ & + a_6[\mathbf{v}_2\mathbf{v}_0] + a_7[\mathbf{v}_3\mathbf{v}_6] + a_8[\mathbf{v}_6\mathbf{v}_5] + a_9[\mathbf{v}_4\mathbf{v}_8] + a_{10}[\mathbf{v}_8\mathbf{v}_7], \quad \text{for some } a_i \in \mathbf{Z}. \end{aligned}$$

We leave this as an exercise for the reader. For example, one can start by first replacing any appearing $[\mathbf{v}_2\mathbf{v}_3]$ in z by $[\mathbf{v}_2\mathbf{v}_6] + [\mathbf{v}_6\mathbf{v}_3]$, then replacing any $[\mathbf{v}_2\mathbf{v}_6]$ by $[\mathbf{v}_2\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_6]$, and so on. Each of these replacements produces a new chain that is homologous to the previous one, so that the final chain $z_1$ is homologous to z. This is similar to what we did in Example 7.2.1.6. If we were to start with a 1-cycle z, then $z_1$ will also be a 1-cycle. However, for an element like $z_1$ to be a 1-cycle, it must satisfy two other properties. First, $a_i$ must be 0 for i = 7, 8, 9, 10; otherwise, one or more of the vertices $\mathbf{v}_5$, $\mathbf{v}_6$, $\mathbf{v}_7$, and $\mathbf{v}_8$ would appear in $\partial_1(z_1)$ with a nonzero coefficient. Second, $a_1 = a_2 = a_3$ and $a_4 = a_5 = a_6$; otherwise, $\partial_1(z_1)$ would not be zero. This shows that any 1-cycle z is homologous to a 1-cycle of the form au + bv, where $a, b \in \mathbf{Z}$ and Claim 3 is proved.

**Claim 4.** The homology classes [u] and [v] are linearly independent.

Assume that

$$a[u] + b[v] = 0$$

in $H_1(K)$ for some integers a and b. It follows that the cycle au + bv is a boundary, that is,

$$au + bv = \partial_2(x)$$

for some 2-chain x. Let $[\boldsymbol{\sigma}]$ be any oriented 2-simplex that appears in x. The simplex $\boldsymbol{\sigma}$ will have at least one edge $\mathbf{v}_i\mathbf{v}_j$ that is different from those appearing in u and v. Since every edge belongs to precisely two 2-simplices, let $\boldsymbol{\sigma}'$ be the other 2-simplex that has $\mathbf{v}_i\mathbf{v}_j$ for an edge. The only way that the coefficient of $[\mathbf{v}_i\mathbf{v}_j]$ will vanish in $\partial_2(x)$ is the coefficients of $[\boldsymbol{\sigma}]$ and $[\boldsymbol{\sigma}']$ in x are equal. A simple extension of this argument shows that every oriented 2-simplex $[\boldsymbol{\sigma}]$ of K must appear in x with the same coefficient. Therefore, $x = k\Sigma$, for some integer k and where $\Sigma$ is the 2-chain defined earlier, and

$$\partial_2(x) = \partial_2(k\Sigma) = k\partial_2(\Sigma) = k \cdot 0 = 0.$$

This means that a = b = 0 and Claim 4 is proved.

Claim 3 and 4 prove that the map that sends $(a,b) \in \mathbf{Z} \oplus \mathbf{Z}$ to $a[u] + b[v] \in H_1(K)$ is an isomorphism and

$$H_1(K) \approx \mathbf{Z} \oplus \mathbf{Z}.$$

Finally, one can show that

$$H_0(K) \approx \mathbf{Z}$$

using the same argument as in Examples 7.2.1.5 and 7.2.1.6. This finishes Example 7.2.1.7.

**7.2.1.8. Example.** The simplicial complex K in Figure 7.6 triangulates the projective plane. We want to compute the homology groups of K.

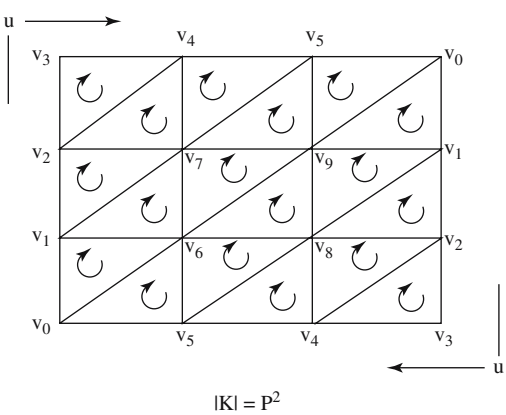



**Figure 7.6.** A triangulating complex for the projective plane.

**Solution.** We shall only sketch the computation of $H_1(K)$ and $H_2(K)$ in this example and leave the details and the computation of the other groups as an exercise to the reader.

**Claim 1.** $Z_2(K) = 0$.

Orient the 2-simplices in K as shown in Figure 7.6. If z is any nonzero 2-cycle, then z must contain all the oriented 2-simplices of K with equal multiplicity. For example, if

$$z = a[\mathbf{v}_6\mathbf{v}_7\mathbf{v}_9] + \ldots,$$

then all the oriented 2-simplices adjacent to $\mathbf{v}_6\mathbf{v}_7\mathbf{v}_9$ must appear with the same multiplicity because that is the only way that the boundary of z can vanish. In other words, z must be of the form $a\Sigma$ where $\Sigma$ is the sum of all the oriented 2-simplices. But

$$\partial(\Sigma) = 2u,$$

where

$$u = [\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_3] + [\mathbf{v}_3\mathbf{v}_4] + [\mathbf{v}_4\mathbf{v}_5] + [\mathbf{v}_5\mathbf{v}_0].$$

It follows that a must be zero and Claim 1 is proved. This also shows that

$$H_2(K) = 0.$$

**Claim 2.** If z is a 1-cycle, then z is homologous to ku for some integer k.

Claim 2 is proved by first showing that z is homologous to a 1-cycle $z_1$ of the form

$$\begin{aligned} z_1 = {} & a_1[\mathbf{v}_0\mathbf{v}_1] + a_2[\mathbf{v}_1\mathbf{v}_2] + a_3[\mathbf{v}_2\mathbf{v}_3] + a_4[\mathbf{v}_3\mathbf{v}_4] + a_5[\mathbf{v}_4\mathbf{v}_5] \\ & + a_6[\mathbf{v}_5\mathbf{v}_0] + a_7[\mathbf{v}_4\mathbf{v}_7] + a_8[\mathbf{v}_7\mathbf{v}_6] + a_9[\mathbf{v}_5\mathbf{v}_9] + a_{10}[\mathbf{v}_9\mathbf{v}_8], \quad \text{for some } a_i \in \mathbf{Z}. \end{aligned}$$

But since $\partial_1(z_1) = 0$ we must have $a_7 = a_8 = a_9 = a_{10} = 0$. Finally, one shows that all the remaining 1-simplices in $z_1$ must appear with the same multiplicity, proving Claim 2.

**Claim 3.** $[u] \neq 0$.

If $u = \partial_2(c)$ for some 2-chain c, then one can show by an argument similar to the one in the proof of Claim 1 that all the oriented 2-simplices must appear in c with multiplicity 1, that is, $c = \Sigma$. This contradicts the fact that $\partial(\Sigma) = 2u$, and proves Claim 3.

Claims 2 and 3 and the fact that $2[u] = 0$ (since $\partial(\Sigma) = 2u$) clearly prove that

$$H_1(K) \approx \mathbf{Z}_2.$$

This finishes what we have to say about Example 7.2.1.8.

One can see from the examples above that homology groups of a simplicial complex K do give us important information about the holes in the underlying space $|K|$. A point, which has no holes at all, had zero homology groups above dimension zero. In Examples 7.2.1.5 and 7.2.1.7 the generators of $H_1(K)$ were associated in a natural way with one-dimensional holes. The two-dimensional hole in the sphere and the torus corresponded to the generator of $H_2(K)$. The fact that the group $H_0(K)$ was nonzero in all of our examples (it was isomorphic to $\mathbf{Z}$) also makes sense if one recalls that 0-dimensional holes intuitively correspond to imbedded 0-spheres and $\mathbf{S}^0$ consists of two points. Exercise 7.2.1.2 asks you to prove that the rank of $H_0(K)$ equals the number of connected components of $|K|$. The spaces in our examples were all connected and had a single connected component.

Working through our examples should also have brought out another point, namely, the **algebraic** nature of homology theory. Although our original motivation was to detect **geometric** "spherical" holes (perhaps even imbedded spheres), the "homological holes," or k-cycles, are more general. For example, the results in the case of the projective plane may have been somewhat unexpected for someone new to homology (twice the generator of $H_1(K)$ was zero and there was no 2-cycle) but they all make sense once one understands this algebraic nature of homology a little better. A k-cycle for a simplicial complex cannot always be represented as an imbedded k-sphere in $|K|$ and an imbedded k-sphere in $|K|$ whose corresponding k-cycle is homologous to zero does not necessarily bound a $(k+1)$-disk in $|K|$. The general question of when k-cycles can be represented by imbedded k-spheres and when imbedded k-spheres bound $(k+1)$-disks is extremely interesting but often difficult to answer, even for manifolds. For example, it is already nontrivial to determine those 1-cycles that can be represented by imbedded circles in the simple case of the torus. The fact is that homology theory is really associated to **abstract** simplicial complexes because all that it needs is an appropriate operator $\partial_q$ on linear combinations of formal symbols of the form $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$. We may have used some geometric intuition to motivate our proofs, but the proofs themselves were independent of it. This is the reason that the study of topology that deals with simplicial complexes is called combinatorial topology. We shall see in Section 7.2.5 how a computer can compute homology groups. However we look at it though, we should be excited by the prospect that we have a theory that detects geometric invariants.

We move on and introduce some more standard terminology. Let K be a simplicial complex. The group $C_q(K)$ is by definition a finitely generated free abelian group with the q-simplices of K forming a set of generators. Since subgroups and quotient groups of finitely generated abelian groups are again finitely generated, we conclude that $H_q(K)$ is finitely generated. It follows from the fundamental theorem about such groups (Theorem B.5.7) that

$$H_q(K) \approx F_q \oplus T_q$$

where $F_q$ is a free group and $T_q$ is the torsion subgroup of $H_q(K)$.

**Definition.** The rank of $F_q$, which is also the rank of $H_q(K)$, is called the *qth Betti number* of K and will be denoted by $\beta_q(K)$. The torsion coefficients of $T_q$ are called the *qth torsion coefficients* of K.

Clearly, once one knows the Betti numbers and torsion coefficients, one knows the homology groups. The Betti numbers and torsion coefficients for Examples 7.2.1.4–8 above are easily determined.

Let us show how homology groups can be used to tackle the problem of classifying topological spaces. At the moment homology groups have only been defined for simplicial complexes, but we could define the homology groups of a polyhedron to be the homology groups of a simplicial complex that triangulates it. The problem is that spaces can be triangulated in many different ways. Therefore, the next theorem, which asserts the topological invariance of homology groups, is essential here. Its proof will be given shortly.

**7.2.1.9. Theorem.** Suppose that K and L are simplicial complexes and that $|K|$ is homeomorphic to $|L|$. Then $H_q(K)$ is isomorphic to $H_q(L)$ for all q.

**Proof.** This is an immediate corollary of Theorem 7.2.3.1 in Section 7.2.3 since homeomorphisms are homotopy equivalences.

**Definition.** Let **X** be a polyhedron. Choose any triangulation $(K,\varphi)$ for **X** and define the *qth homology group* of **X**, $H_q(\mathbf{X})$, by $H_q(\mathbf{X}) = H_q(K)$. The *qth Betti number* of **X**, $\beta_q(\mathbf{X})$, and the *qth torsion coefficients* of **X** are defined to be the rank and torsion coefficients of K, respectively.

Although the groups $H_q(\mathbf{X})$ are not uniquely defined since a polyhedron has many triangulations, they are well defined up to isomorphism by Theorem 7.2.1.9. This is all we need to be able to conclude that homology groups give us topological invariants of spaces and not just properties of particular triangulations and are therefore exactly the type of computable algebraic invariants we were looking for in Section 6.2. See [AgoM76] for a wide range of applications justifying the hard work we put into our effort. Some of these will also be discussed in Section 7.2.3. The next chapter will also look into the subject further, but at this point we are lacking one important ingredient of the theory, namely, we do not yet know how homology groups behave with respect to maps.

Table 7.2.1.1 summarizes the computations we have made so far and some that come from Exercises 7.2.1.3–4. The results are stated for polyhedra rather than simplicial complexes because that is what we are after anyway, not the intermediate structures, the simplicial complexes.

We finish this section with a comment about the concept of orientability. Looking at Table 7.2.1.1 we see that what distinguishes closed compact orientable surfaces **S** from nonorientable ones is the group $H_2(\mathbf{S})$, or, to put it another way, the ability to find a nonzero 2-cycle. A little reflection shows that we got a nonzero 2-cycle precisely because it was possible to orient the 2-simplices of a triangulating simplicial complex so that adjacent 2-simplices induced the opposite orientation on the 1-simplex in their boundary which they had in common. This shows that the condition of orientability as described in Chapter 1 is actually easily computed by combinatorial methods – one simply needs to find a 2-cycle.

## 7.2.2 Induced Maps

In Section 7.2.1 we defined a mapping from simplicial complexes K to their homology groups $H_q(K)$. We now show how this mapping of **objects** extends to a mapping of **maps** in a natural way.

**Table 7.2.1.1** Some spaces and their homology groups.

| $\mathbf{X}$ | $H_0(\mathbf{X})$ | $H_1(\mathbf{X})$ | $H_2(\mathbf{X})$ $i > 2$ | $H_i(\mathbf{X})$ |
|---|---|---|---|---|
| $\mathbf{D}^0$ | $\mathbf{Z}$ | 0 | 0 | 0 |
| $\mathbf{S}^0$ | $\mathbf{Z} \oplus \mathbf{Z}$ | 0 | 0 | 0 |
| $\mathbf{S}^1$ | $\mathbf{Z}$ | $\mathbf{Z}$ | 0 | 0 |
| $\mathbf{S}^2$ | $\mathbf{Z}$ | 0 | $\mathbf{Z}$ | 0 |
| $\mathbf{S}^1 \times \mathbf{S}^1$ | $\mathbf{Z}$ | $\mathbf{Z} \oplus \mathbf{Z}$ | $\mathbf{Z}$ | 0 |
| $\mathbf{P}^2$ | $\mathbf{Z}$ | $\mathbf{Z}_2$ | 0 | 0 |
| The Klein bottle | $\mathbf{Z}$ | $\mathbf{Z} \oplus \mathbf{Z}_2$ | 0 | 0 |
| Orientable surface (genus k) | $\mathbf{Z}$ | $k\,(\mathbf{Z} \oplus \mathbf{Z})$ | $\mathbf{Z}$ | 0 |
| Nonorientable surface (genus k) | $\mathbf{Z}$ | $\mathbf{Z}^{k-1} \oplus \mathbf{Z}_2$ | 0 | 0 |

**Definition.** Let K and L be simplicial complexes and let $f: K \to L$ be a simplicial map. Define maps

$$f_{\#q}: C_q(K) \to C_q(L)$$

as follows:

If $q < 0$ or $q > \dim K$, then $f_{\#q} = 0$.
If $0 \le q \le \dim K$, then $f_{\#q}$ is the unique homomorphism defined by the condition that

$$\begin{aligned} f_{\#q}([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) &= [f(\mathbf{v}_0)f(\mathbf{v}_1)\cdots f(\mathbf{v}_q)], \quad \text{if } f(\mathbf{v}_i) \ne f(\mathbf{v}_j) \quad \text{for } i \ne j, \\ &= 0, \quad \text{otherwise,} \end{aligned}$$

for each oriented q-simplex $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$ of $C_q(K)$. (The map $f_{\#q}$ is well defined because the group $C_q(K)$ is a free group with generators $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$.)

**7.2.2.1. Lemma.** $\partial_q \circ f_{\#q} = f_{\#q-1} \circ \partial_q$ for all q. In other words, for all q there is a commutative diagram

$$\begin{array}{ccc} C_q(K) & \xrightarrow{f_{\#q}} & C_q(L) \\ \partial_q \downarrow & & \downarrow \partial_q \\ C_{q-1}(K) & \xrightarrow[f_{\#q-1}]{} & C_{q-1}(L) \end{array} .$$

**Proof.** Clearly, it suffices to show that

$$(\partial_q \circ f_{\#q})([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) = (f_{\#q-1} \circ \partial_q)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q])$$

for every oriented q-simplex $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]$ in K and we do this by computing both sides of this equation.

**Case 1.** The vertices $f(\mathbf{v}_0)$, $f(\mathbf{v}_1)$, . . . , and $f(\mathbf{v}_q)$ are all distinct.

In this case,

$$\begin{aligned}(\partial_q \circ f_{\#q})([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) &= \partial_q([f(\mathbf{v}_0)f(\mathbf{v}_1)\cdots f(\mathbf{v}_q)])\\ &= \sum_{i=0}^{q}(-1)^i[f(\mathbf{v}_0)\cdots \widehat{f(\mathbf{v}_i)}\cdots f(\mathbf{v}_q)]\\ &= \sum_{i=0}^{q}(-1)^i f_{\#q-1}[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\\ &= f_{\#q-1}\left(\sum_{i=0}^{q}(-1)^i[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\right)\\ &= (f_{\#q-1}\circ\partial_q)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]).\end{aligned}$$

**Case 2.** The vertices $f(\mathbf{v}_1)$, $f(\mathbf{v}_2)$, . . . , and $f(\mathbf{v}_q)$ are all distinct, but $f(\mathbf{v}_0) = f(\mathbf{v}_1)$.

The assumption $f(\mathbf{v}_0) = f(\mathbf{v}_1)$ implies that

$$(\partial_q \circ f_{\#q})([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) = \partial_q(0) = 0$$

and

$$\begin{aligned}(f_{\#q-1}\circ\partial_q)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) &= (f_\#)_{q-1}\left(\sum_{i=0}^{q}(-1)^i[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\right)\\ &= \sum_{i=0}^{q}(-1)^i(f_\#)_{q-1}[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q]\\ &= (-1)^0[f(\mathbf{v}_1)f(\mathbf{v}_2)\cdots f(\mathbf{v}_q)] + (-1)^1[f(\mathbf{v}_0)f(\mathbf{v}_2)\cdots f(\mathbf{v}_q)]\\ &\qquad + (-1)^2\cdot 0 + \cdots + (-1)^q\cdot 0\\ &= [f(\mathbf{v}_1)f(\mathbf{v}_2)\cdots f(\mathbf{v}_q)] - [f(\mathbf{v}_0)f(\mathbf{v}_2)\cdots f(\mathbf{v}_q)]\\ &= 0.\end{aligned}$$

**Case 3.** The vertices $f(\mathbf{v}_0)$, $f(\mathbf{v}_1)$, . . . , $\widehat{f(\mathbf{v}_i)}$, . . . , $f(\mathbf{v}_q)$ are all distinct, but $f(\mathbf{v}_i) = f(\mathbf{v}_j)$ for some $i < j$.

This case follows easily from Case 2 because $[\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q] = \pm[\mathbf{v}_i\mathbf{v}_j\mathbf{v}_0\mathbf{v}_1 \cdots \hat{\mathbf{v}}_i \cdots \hat{\mathbf{v}}_j \cdots \mathbf{v}_q]$.

**Case 4.** There exist distinct indices i, j, and k, such that $f(\mathbf{v}_i) = f(\mathbf{v}_j) = f(\mathbf{v}_k)$.

In this case,

$$(\partial_q \circ f_{\#q})([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) = \partial_q(0) = 0$$

and

$$((f_\#)_{q-1} \circ \partial_q)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q]) = (f_\#)_{q-1}\left(\sum_{t=0}^{q}(-1)^t[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_t \cdots \mathbf{v}_q]\right)$$
$$= \sum_{t=0}^{q}(-1)^t(f_\#)_{q-1}[\mathbf{v}_0 \cdots \hat{\mathbf{v}}_t \cdots \mathbf{v}_q] = 0.$$

The lemma is proved.

Now the maps $f_{\#q}$ are no more interesting by themselves than were the chain groups $C_q(K)$. What will be important are the maps that they induce on the homology groups, and Lemma 7.2.2.1 is essential for that. We generalize the construction somewhat.

**Definition.** A *chain map*

$$\varphi: C_\#(K) \to C_\#(L)$$

is a "vector" $\varphi = (\ldots,\varphi_{-1},\varphi_0,\varphi_1,\ldots)$ of homomorphisms $\varphi_q: C_q(K) \to C_q(L)$ satisfying $\partial_q \circ \varphi_q = \varphi_{q-1} \circ \partial_q$.

Having a chain map $(\ldots,\varphi_{-1},\varphi_0,\varphi_1,\ldots)$ is equivalent to having a commutative diagram

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{\partial_{q+2}} & C_{q+1}(K) & \xrightarrow{\partial_{q+1}} & C_q(K) & \xrightarrow{\partial_q} & C_{q-1}(K) & \to & \cdots \\
 & & \downarrow\varphi_{q+1} & & \downarrow\varphi_q & & \downarrow\varphi_{q-1} & & \\
\cdots & \xrightarrow[\partial_{q+2}]{} & C_{q+1}(L) & \xrightarrow[\partial_{q+1}]{} & C_q(L) & \xrightarrow[\partial_q]{} & C_{q-1}(L) & \to & \cdots
\end{array}$$

Note that $f_\# = (\ldots, f_{\#-1},f_{\#0},f_{\#1},\ldots)$ is a chain map by Lemma 7.2.2.1 called the *chain map induced by the simplicial map f*.

**7.2.2.2. Lemma.** Let $\varphi: C_\#(K) \to C_\#(L)$ be an arbitrary chain map. Then

(1) $\varphi(Z_q(K)) \subseteq Z_q(L)$, for all q.
(2) $\varphi(B_q(K)) \subseteq B_q(L)$, for all q.

**Proof.** This follows easily from the definition of a chain map. If $z \in Z_q(K)$, then

$$\partial_q(\varphi_q(z)) = \varphi_{q-1}(\partial_q(z)) = \varphi_{q-1}(0) = 0,$$

which proves (1). To prove (2), note that

$$\varphi_q(\partial_{p+1}(c)) = \partial_{p+1}(\varphi_{q+1}(c)).$$

**Definition.** Let $\varphi: C_\#(K) \to C_\#(L)$ be an arbitrary chain map. Define maps

$$\varphi_{*q}: H_q(K) \to H_q(L)$$

by

$$\varphi_{*q}([z]) = [\varphi_q(z)], \quad z \in Z_q(K).$$

**7.2.2.3. Lemma.** $\varphi_{*q}$ is a well-defined homomorphism.

**Proof.** First of all, by Lemma 7.2.2.2(1), the definition makes sense, since $\varphi_q(z) \in Z_q(L)$. To show that $\varphi_{*q}$ is well defined, let $a \in H_q(K)$ and assume that $a = [z] = [z']$, $z, z' \in Z_q(K)$. Then $z - z'$ belongs to $B_q(K)$. Therefore,

$$\varphi_q(z) - \varphi_q(z') = \varphi_q(z - z') \in B_q(L)$$

by Lemma 7.2.2.2(2), that is, $[\varphi_q(z)] = [\varphi_q(z')]$. This proves that $\varphi_{*q}$ is well defined.

Next, let $[z_i] = z_i + B_q(K)$ be elements of $H_q(K)$. Then

$$\begin{aligned}
\varphi_{*q}([z_1] + [z_2]) &= \varphi_{*q}((z_1 + z_2) + B_q(K)) \\
&= \varphi_q(z_1 + z_2) + B_q(L) \\
&= (\varphi_q(z_1) + \varphi_q(z_2)) + B_q(L) \\
&= (\varphi_q(z_1) + B_q(L)) + (\varphi_q(z_2) + B_q(L)) \\
&= \varphi_{*q}([z_1]) + \varphi_{*q}([z_2]).
\end{aligned}$$

Thus, $\varphi_{*q}$ is a homomorphism and Lemma 7.2.2.3 is proved.

**Definition.** The maps $\varphi_{*q}$ are called the *homomorphisms on homology induced by the chain map* $\varphi$. In particular, if $f: K \to L$ is a simplicial map, we shall let

$$f_{*q}: H_q(K) \to H_q(L)$$

denote the map on the homology group induced by the chain map $f_\#$.

Consider the simplicial complex $K = \partial \langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 \rangle$. The next two examples compute $f_{*q}$ for two simplicial maps $f: K \to K$.

**7.2.2.4. Example.** To compute $f_{*q}$ when f is the constant map defined by $f(\mathbf{v}_i) = \mathbf{v}_0$.

**Solution.** The given f induces the constant map $|f| : |K| \to |K|$, $|f|(x) = \mathbf{v}_0$. We know from Example 7.2.1.5 that

$$H_0(K) \approx H_1(K) \approx \mathbf{Z}$$

and

$$H_q(K) = 0 \quad \text{for} \quad q > 1,$$

so that we only have to worry about what happens in dimensions 0 and 1. The map $f_{\#1}: C_1(K) \to C_1(K)$ is obviously the zero map by definition, and so $f_{*1}: H_1(K) \to H_1(K)$

is also the zero map. Next, note that the group $H_0(K)$ is generated by the element $\mathbf{v}_0 + B_0(K)$ and

$$f_{*0}(\mathbf{v}_0 + B_0(K)) = f_{\#0}(\mathbf{v}_0) + B_0(K) = \mathbf{v}_0 + B_0(K).$$

This implies that $f_{*0}$ is the identity map.

**7.2.2.5. Example.** To compute $f_{*q}$ when f is defined by $f(\mathbf{v}_0) = \mathbf{v}_0$, $f(\mathbf{v}_1) = \mathbf{v}_2$, and $f(\mathbf{v}_2) = \mathbf{v}_1$.

**Solution.** It follows from an argument similar to the one in the previous example that $f_{*q} = 0$ if $q > 1$ and $f_{*0}$ is the identity map. Only $f_{*1}$ is different this time. Recall from Example 7.2.1.5 that

$$a = ([\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_0]) + B_1(K)$$

is a generator of $H_1(K)$. Since

$$\begin{aligned} f_{*1}(a) &= f_{\#1}([\mathbf{v}_0\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_0]) + B_1(K) \\ &= ([f(\mathbf{v}_0)f(\mathbf{v}_1)] + [f(\mathbf{v}_1)f(\mathbf{v}_2)] + [f(\mathbf{v}_2)f(\mathbf{v}_0)]) + B_1(K) \\ &= ([\mathbf{v}_0\mathbf{v}_2] + [\mathbf{v}_2\mathbf{v}_1] + [\mathbf{v}_1\mathbf{v}_0]) + B_1(K) \\ &= -a \end{aligned}$$

it follows that $f_{*1}$ is the negative of the identity map.

The next lemma lists some basic properties of the maps $f_{\#q}$ and $f_{*q}$ that are easy to prove.

**7.2.2.6. Lemma.** Let $f: K \to L$ and $g: L \to M$ be simplicial maps between simplicial complexes. Then

(1) $(g \circ f)_{\#q} = g_{\#q} \circ f_{\#q}: C_q(K) \to C_q(M)$.
(2) $(g \circ f)_{*q} = g_{*q} \circ f_{*q}: H_q(K) \to H_q(M)$.
(3) If $K = L$ and $f = 1_K$, then $f_{\#q}$ and $f_{*q}$ are also the identity homomorphisms.

**Proof.** This is Exercise 7.2.2.1.

Now simplicial complexes and maps are basically only tools for studying topological spaces and continuous maps. We shall show next how continuous maps induce homomorphisms on homology groups.

**Definition.** Let K and L be simplicial complexes and suppose that $f: |K| \to |L|$ is a continuous map. A *simplicial approximation to f* is a simplicial map $\varphi: K \to L$ with the following property: If $\mathbf{x} \in |K|$ and if $f(\mathbf{x}) \in \boldsymbol{\sigma}$ for some simplex $\boldsymbol{\sigma} \in L$, then $|\varphi|(\mathbf{x}) \in \boldsymbol{\sigma}$.

The next lemma summarizes two important properties of simplicial approximations.

**7.2.2.7. Lemma.** Let $f: |K| \to |L|$ be a continuous map and suppose that $\varphi: K \to L$ is a simplicial approximation to f.

(1) The map $|\varphi|: |K| \to |L|$ is homotopic to f.
(2) If $f = |\psi|$, where $\psi: K \to L$ is a simplicial map, then $\psi = \varphi$.

**Proof.** To prove (1), define a homotopy $h: |K| \times [0,1] \to |L|$ between f and $|\varphi|$ by

$$h(\mathbf{x}, t) = t\,|\varphi|(\mathbf{x}) + (1-t)f(\mathbf{x})$$

for $\mathbf{x} \in |K|$ and $t \in [0,1]$. That $h(\mathbf{x},t)$ actually lies in $|L|$ follows from the fact that $|\varphi|(\mathbf{x})$ and $f(\mathbf{x})$ lie in a simplex of L, which means that the segment $[|\varphi|(\mathbf{x}), f(\mathbf{x})]$ is contained in $|L|$ because simplices are convex.

To prove (2), let $\mathbf{v}$ be a vertex of K. Then $\mathbf{w} = f(\mathbf{v})$ is a vertex of L. Since a vertex is also a 0-simplex, the definition of a simplicial approximation implies that $\varphi(\mathbf{v}) = \mathbf{w}$. This proves the lemma.

Part (2) of Lemma 7.2.2.7 means that the only simplicial approximation to a simplicial map is the map itself. An arbitrary continuous map does **not** have a unique simplicial approximation, however.

If K is a simplicial complex, define a new simplicial complex, denoted by sd(K), as follows:

(1) The vertices of sd(K) are the barycenters $b(\boldsymbol{\sigma})$ of the simplices $\boldsymbol{\sigma}$ in K.
(2) The q-simplices of sd(K), $q > 0$, are all the q-simplices of the form $b(\boldsymbol{\sigma}_0)b(\boldsymbol{\sigma}_1) \cdots b(\boldsymbol{\sigma}_q)$, where the $\boldsymbol{\sigma}_i$ are distinct simplices of K and $\boldsymbol{\sigma}_0 \prec \boldsymbol{\sigma}_1 \prec \ldots \prec \boldsymbol{\sigma}_q$.

It is easy to show that sd(K) is a simplicial complex (Exercise 7.2.2.3) that is a subdivision of K. Clearly, $|sd(K)| = |K|$.

**Definition.** The simplicial complex sd(K) is called the *(first) barycentric subdivision of K*. The *nth barycentric subdivision of K*, denoted by $sd^n(K)$, is defined inductively by

$$sd^0(K) = K,$$
$$sd^n(K) = sd(sd^{n-1}(K)), \quad \text{for } n \geq 1.$$

Figure 7.7 shows a simplex and its barycentric subdivision.

**7.2.2.8. Theorem.** (The Simplicial Approximation Theorem) Let K and L be simplicial complexes and suppose that $f: |K| \to |L|$ is a continuous map. Then there is an integer $N \geq 0$ such that for each $n \geq N$, f admits a simplicial approximation $\varphi: sd^n(K) \to L$.

**Proof.** See [AgoM76].

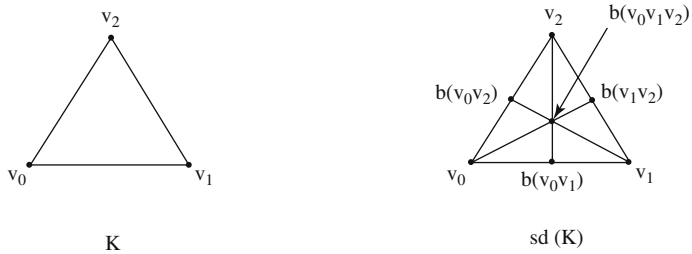



**Figure 7.7.** A barycentric subdivision.

Associated to barycentric subdivisions are natural homomorphisms

$$\mathrm{sd}_{\#q} : C_q(K) \to C_q(\mathrm{sd}(K))$$

that correspond to sending an oriented simplex $[\boldsymbol{\sigma}]$ to the sum of the oriented simplices into which the barycentric subdivision divides $\boldsymbol{\sigma}$. For example,

$$\begin{aligned}\mathrm{sd}_{\#2}([\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2]) = {} & [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathbf{v}_0\mathrm{b}(\mathbf{v}_0\mathbf{v}_1)] + [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathrm{b}(\mathbf{v}_0\mathbf{v}_1)\mathbf{v}_1] + [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathbf{v}_1\mathrm{b}(\mathbf{v}_1\mathbf{v}_2)] \\ & + [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathrm{b}(\mathbf{v}_1\mathbf{v}_2)\mathbf{v}_2] + [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathbf{v}_2\mathrm{b}(\mathbf{v}_0\mathbf{v}_2)] + [\mathrm{b}(\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2)\mathrm{b}(\mathbf{v}_0\mathbf{v}_2)\mathbf{v}_0].\end{aligned}$$

See Figure 7.7. More precisely, define the maps $\mathrm{sd}_{\#q}$ inductively on the oriented simplices as follows:

(1) If $\mathbf{v}$ is a vertex of K, then $\mathrm{sd}_{\#0}(\mathbf{v}) = \mathbf{v}$.
(2) Assume $0 < q < \dim K$ and $\mathrm{sd}_{\#q-1}$ has been defined. If $[\boldsymbol{\sigma}]$ is an oriented q-simplex of K, then

$$\mathrm{sd}_{\#q}([\boldsymbol{\sigma}]) = \mathrm{b}(\boldsymbol{\sigma})\mathrm{sd}_{\#q-1}(\partial_q([\boldsymbol{\sigma}])).$$

(We are using the expression $\mathbf{w}[\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]$ to denote the oriented simplex $[\mathbf{w}\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_q]$ and let this operation distribute over sums.)

If $q < 0$ or $\dim K < q$, then we define $\mathrm{sd}_{\#q}$ to be the zero map.

**7.2.2.9. Lemma.** The maps $\mathrm{sd}_{\#q}$ are well-defined homomorphisms. Furthermore, $\partial_q \circ \mathrm{sd}_{\#q} = \mathrm{sd}_{\#q-1} \circ \partial_q$, so that $\mathrm{sd}_{\#} = (\ldots, \mathrm{sd}_{\#-1}, \mathrm{sd}_{\#0}, \mathrm{sd}_{\#1}, \ldots)$ is a chain map that induces homomorphisms

$$\mathrm{sd}_{*q} : H_q(K) \to H_q(\mathrm{sd}(K)).$$

**Proof.** This is an easy exercise. See [AgoM76].

We can extend our definitions and define homomorphisms

$$\mathrm{sd}^n_{\#q} : C_q(K) \to C_q\big(\mathrm{sd}^n(K)\big)$$

inductively by

$$\begin{aligned} sd^n_{\#q} &= \text{zero map, for } n < 0, \\ sd^0_{\#q} &= \text{identity map of } C_0(K), \\ sd^n_{\#q} &= sd_{\#q} \circ sd^{n-1}_{\#q}, \quad \text{for } n > 0. \end{aligned}$$

The maps $sd^n_{\#q}$ induce homomorphisms (actually isomorphisms)

$$sd^n_{*q}: \ H_q(K) \to H_q\big(sd^n(K)\big).$$

We are now ready to show how continuous maps induce homomorphisms on homology groups. Let K and L be simplicial complexes and let

$$f: \ |K| \to |L|$$

be a continuous map. The Simplicial Approximation Theorem implies that there is an $n \geq 0$, such that f admits a simplicial approximation

$$\varphi: \ sd^n(K) \to L.$$

**Definition.** The homomorphism

$$f_{*q}: \ H_q(K) \to H_q(L)$$

defined by

$$f_{*q} = \varphi_{*q} \circ sd^n_{*q}$$

is called the *homomorphism induced on the qth homology group* by the continuous map f.

**7.2.2.10. Lemma.**

(1) $f_{*q}$ is a well-defined homomorphism.
(2) If $K = L$ and $f = 1_K$, then $f_{*q}$ is the identity homomorphism.
(3) If M is a simplicial complex and $g: |L| \to |M|$ is a continuous map, then $(g \circ f)_{*q} = g_{*q} \circ f_{*q}$.

**Proof.** See [AgoM76].

**7.2.2.11. Theorem.** Let K and L be simplicial complexes and suppose that f, $g: |K| \to |L|$ are continuous maps that are homotopic. Then $f_{*q} = g_{*q}: H_q(K) \to H_q(L)$ for all q.

**Proof.** See [AgoM76].

### 7.2.3 Applications of Homology Theory

Before describing some applications, it is worthwhile to briefly pause and summarize what we have accomplished so far; otherwise, it is easy to lose sight of the global picture and get lost in a sequence of lemmas and theorems. The main results can be summarized by the following:

**Fact 1.** For every simplicial complex K and every integer q there is an abelian group $H_q(K)$ called the qth homology group of K.

**Fact 2.** For every continuous map $f: |K| \to |L|$ between the underlying spaces of two simplicial complexes K and L there are a homomorphisms

$$f_{*q} : H_q(K) \to H_q(L)$$

whose natural properties are best summarized by the commutative diagram

$$\begin{tikzcd}
 & & |L| \arrow[dr, "g"] \arrow[dd] & \\
|K| \arrow[r, "1_{|K|}"] \arrow[dd] & |K| \arrow[ur, "f,\, f'"] \arrow[dd] \arrow[r, phantom, "f \simeq f'"] & & |M| \arrow[dd] \\
 & & H_q(L) \arrow[dr, "g_{*q}"] & \\
H_q(K) \arrow[r, "1_{H_q(K)} = (1_{|K|})_{*q}"'] & H_q(K) \arrow[ur, "f_{*q} = (f')_{*q}"] \arrow[rr, "(g \circ f)_{*q}"'] & & H_q(M)
\end{tikzcd}$$

The top line in the diagram deals with simplicial complexes and maps and the bottom lines deal with groups and homomorphisms.

For our purposes, Facts 1 and 2 contain essentially everything that we need to know about homology groups and induced maps. Many of our applications will follow in a purely formal way from Facts 1 and 2 with the geometry being irrelevant. Actual definitions are only needed for a few specific computations. There is one caveat though. We would really like to have well-defined homology groups and induced maps associated to polyhedra and their continuous maps. Singular homology theory (see Section 7.6) accomplishes that, but we shall at times pretend that we have this here also. To avoid such pretense and restore rigor we could pick a fixed triangulation for each polyhedron and translate continuous maps between them to maps between the underlying spaces of the simplicial complexes. This would validate our arguments but the messy details would obscure geometric ideas. By the way, homology theory can be described axiomatically by means of the so-called *Eilenberg-Steenrod axioms*. Facts 1 and 2 correspond to a subset of these axioms. Statements made in Chapter 6 about algebraic topology associating "algebraic invariants" to spaces should make a lot more sense now.

**7.2.3.1. Theorem.** Two simplicial complexes K and L with homotopy equivalent underlying spaces have isomorphic homology groups.

**Proof.** Let $f: |K| \to |L|$ and $g: |L| \to |K|$ be continuous maps such that $g \circ f \simeq 1_{|K|}$ and $f \circ g \simeq 1_{|L|}$. Then Fact 2 implies that

$$g_{*q} \circ f_{*q} = 1_{H_q(K)} \quad \text{and} \quad f_{*q} \circ g_{*q} = 1_{H_q(L)}$$

for all q. It follows that $f_{*q}$ is an isomorphism and the theorem is proved.

**7.2.3.2. Corollary.** Homotopy equivalent polyhedra have isomorphic homology groups.

**7.2.3.3. Corollary.** If a polyhedron **X** has the homotopy type of a point, then

$$H_q(\mathbf{X}) \approx \begin{Bmatrix} \mathbf{Z}, & q = 0 \\ 0, & q > 0 \end{Bmatrix}.$$

In particular, these are the homology groups of $\mathbf{D}^n$.

There is one consequence of Theorem 7.2.3.1 that would be disappointing to anyone who might have hoped to use homology groups to classify topological spaces. They are not strong enough invariants to distinguish spaces up to homeomorphism. For example, Corollary 7.2.3.3 shows that both a single point and the disk $\mathbf{D}^n$ have the same homology groups but are clearly not homeomorphic. The best we could hope for now is that they distinguish spaces up to homotopy type. Unfortunately, they fail to do even that except in special cases. (There exist polyhedra, such as the spaces in Example 7.2.4.7, that have isomorphic homology groups but that are **not** homotopy equivalent.) Nevertheless, homology groups are strong enough to enable one to prove many negative results, that is, if one can show that two spaces have nonisomorphic homology groups, then it follows that the are **not** homeomorphic. In fact, they would not even have the same homotopy type.

Before we state several invariance results that can be proved using homology groups, we need to compute the homology groups for the higher-dimensional spheres.

**7.2.3.4. Theorem.** If $n \geq 1$, then

$$H_k(S^{n-1}) \approx \begin{Bmatrix} \mathbf{Z} \oplus \mathbf{Z}, & \text{if} \quad n = 1 \quad \text{and} \quad k = 0 \\ \mathbf{Z}, & \text{if} \quad n \geq 2 \quad \text{and either} \quad k = 0 \quad \text{or} \quad k = n-1 \\ 0, & \text{if} \quad k \neq 0 \quad \text{or} \quad n-1. \end{Bmatrix}$$

**Proof.** The case n = 1 is left as an easy exercise for the reader. Assume that $n \geq 2$. Let $\boldsymbol{\sigma} = \mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_n$ be any n-simplex. Let $M = \langle \boldsymbol{\sigma} \rangle$ and $N = \partial M$ be the simplicial complexes associated to the simplex and its boundary. Since $\mathbf{S}^{n-1}$ is homeomorphic to $\partial\boldsymbol{\sigma} = |N|$, it suffices to compute $H_k(N)$. The definition of the simplicial homology groups implies that

$$\begin{aligned} B_k(N) &= B_k(M), \quad \text{for} \quad 0 \leq k \leq n-2 \quad \text{or} \quad k \geq n, \\ Z_k(N) &= Z_k(M), \quad \text{for all } k. \end{aligned}$$

Therefore,

$$H_k(N) = H_k(M), \quad \text{for} \quad k \neq n-1,$$

and Corollary 7.2.3.3 proves Theorem 7.2.3.4 for these values of k.

To compute $H_{n-1}(N)$, note that

$$(\partial_{n-1} \circ \partial_n)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_n]) = 0,$$

and so

$$\Sigma = \partial_n([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_n]) \in Z_{n-1}(N).$$

We shall show that $\Sigma$ is in fact a generator of $Z_{n-1}(N)$. If

$$z = \sum_{i=0}^{n} a_i[\mathbf{v}_0\mathbf{v}_1 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_n] \in Z_{n-1}(N),$$

then

$$\begin{aligned} 0 = \partial_{n-1}(z) &= \sum_{i=0}^{n} a_i \partial_{n-1}([\mathbf{v}_0\mathbf{v}_1 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_n]) \\ &= \sum_{i=0}^{n} a_i \left( \sum_{j=0}^{i-1} (-1)^j [\mathbf{v}_0\mathbf{v}_1 \cdots \hat{\mathbf{v}}_j \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_n] + \sum_{j=i+1}^{n} (-1)^{j-1} [\mathbf{v}_0\mathbf{v}_1 \cdots \hat{\mathbf{v}}_i \cdots \hat{\mathbf{v}}_j \cdots \mathbf{v}_n] \right). \end{aligned}$$

Let $s < t$. The coefficient of the oriented $(n-2)$-simplex $[\mathbf{v}_0\mathbf{v}_1 \ldots \hat{\mathbf{v}}_s \ldots \hat{\mathbf{v}}_t \ldots \mathbf{v}_n]$ is

$$(-1)^s a_t + (-1)^{t-1} a_s.$$

Since this coefficient has to vanish, it is easy to check that $z = a_0\Sigma$. It follows that $Z_{n-1}(N) = \mathbf{Z}\Sigma$. But $B_{n-1}(N) = 0$ and $Z_{n-1}(N)$ has no elements of finite order, so that

$$H_{n-1}(N) = Z_{n-1}(N) \approx \mathbf{Z},$$

and the theorem is proved.

**7.2.3.5. Theorem.**

(1) The spheres $\mathbf{S}^n$ and $\mathbf{S}^m$ have the same homotopy type only when $n = m$. In particular, $\mathbf{S}^n$ is homeomorphic to $\mathbf{S}^m$ if and only if $n = m$.
(2) The Euclidean space $\mathbf{R}^n$ is homeomorphic to $\mathbf{R}^m$ only when $n = m$.

**Proof.** Part (1) follows from Theorem 7.2.3.4 and Corollary 7.2.3.2. To prove part (2), we use the stereographic projection

$$p_n : \mathbf{S}^n - \mathbf{e}_{n+1} \to \mathbf{R}^n.$$

Suppose that $h: \mathbf{R}^n \to \mathbf{R}^m$ is a homeomorphism. Define

$$H: \mathbf{S}^n \to \mathbf{S}^m$$

by

$$H(\mathbf{x}) = \left(p_m^{-1} \circ h \circ p_n\right)(\mathbf{x}), \quad \text{if} \quad \mathbf{x} \neq \mathbf{e}_{n+1}, \quad \text{and}$$
$$H(\mathbf{e}_{n+1}) = \mathbf{e}_{m+1}.$$

The map H will be a homeomorphism and therefore n = m by part (1). The theorem is proved.

The next three theorems are less trivial.

**7.2.3.6. Theorem.** (Invariance of Dimension) If K and L are simplicial complexes with $|K| \approx |L|$, then dim K = dim L.

**Proof.** See [AgoM76].

**Definition.** The *dimension* of a polyhedron is defined to be the dimension of any simplicial complex that triangulates it.

Theorem 7.2.3.6 shows the dimension of a polyhedron is a well-defined topological invariant.

**7.2.3.7. Theorem.** (Invariance of Boundary) If K and L are simplicial complexes and $h: |K| \to |L|$ is a homeomorphism, then $h(|\partial K|) = |\partial L|$.

**Proof.** See [AgoM76].

Theorem 7.2.3.7 makes it possible to define the boundary of a polyhedron.

**Definition.** Let **X** be a polyhedron. Define the *boundary* of **X**, denoted by $\partial\mathbf{X}$, by

$$\partial\mathbf{X} = \varphi(|\partial K|),$$

where $(K,\varphi)$ is any triangulation of **X**.

**7.2.3.8. Theorem** (Invariance of Domain) If **U** and **V** are homeomorphic subsets of $\mathbf{R}^n$ and if **U** is open in $\mathbf{R}^n$, then so is **V**.

**Proof.** See [AgoM76].

Returning to our definition of topological manifolds in Section 5.3, we are finally able to prove the claimed invariance of two aspects of the definition.

**7.2.3.9. Corollary.** The dimension of a topological manifold and its boundary are well defined.

**Proof.** The corollary is an easy consequence of Theorems 7.2.3.5 and 7.2.3.8.

Next, we return to the Euler characteristic as defined in Chapter 6. We are now in a position to put this invariant in a more general context. What we had in Chapter 6 was a combinatorial concept defined for surfaces that was easy to compute by some simple counting and yet was claimed to be a topological invariant. We can now define that topological invariant in a rigorous manner.

**Definition.** If K is a simplicial complex, let $n_q(K)$ denote the number of q-simplices in K and define the *Euler-Poincaré characteristic* of K, $\chi(K)$, by

$$\chi(K) = \sum_{q=0}^{\dim K} (-1)^q n_q(K).$$

What makes $\chi(K)$ a topological invariant is the fact that it is related to the Betti numbers $\beta_q(K)$ of K.

**7.2.3.10. Theorem.** (The Euler-Poincaré Formula) Let K be a simplicial complex. Then

$$\chi(K) = \sum_{q=0}^{\dim K} (-1)^q \beta_q(K).$$

**Proof.** By definition, the boundary map $\partial_q : C_q(K) \to B_{q-1}(K)$ is onto and has kernel $Z_q(K)$ and the natural projection $Z_q(K) \to H_q(K)$ is onto and has kernel $B_q(K)$. Therefore, Theorem B.5.8 implies that

$$\begin{aligned} &\operatorname{rank}(C_q(K)) = \operatorname{rank}(B_{q-1}(K)) + \operatorname{rank}(Z_q(K)) \quad \text{and} \\ &\operatorname{rank}(Z_q(K)) = \operatorname{rank}(H_q(K)) + \operatorname{rank}(B_q(K)). \end{aligned}$$

These identities and the fact that $\operatorname{rank}(C_q(K)) = n_q(K)$ gives us that

$$\begin{aligned} \chi(K) &= \sum_{q=0}^{\dim K} (-1)^q n_q(K) \\ &= \sum_{q=0}^{\dim K} (-1)^q \operatorname{rank}(C_q(K)) \\ &= \sum_{q=0}^{\dim K} (-1)^q [\operatorname{rank}(B_{q-1}(K)) + \operatorname{rank}(H_q(K)) + \operatorname{rank}(B_q(K))] \\ &= \sum_{q=0}^{\dim K} (-1)^q \operatorname{rank}(H_q(K)) + \sum_{q=0}^{\dim K} (-1)^q [\operatorname{rank}(B_{q-1}(K)) + \operatorname{rank}(B_q(K))] \\ &= \sum_{q=0}^{\dim K} (-1)^q \operatorname{rank}(H_q(K)) \end{aligned}$$

since the rank $(B_i(K))$ terms cancel each other in the sum. This proves the theorem because $\beta_q(K) = \text{rank}(H_q(K))$.

It follows from Theorem 7.2.3.10 that this combinatorially defined number $\chi(K)$ is a topological invariant associated to the underlying space $|K|$. In fact, it is more than that and actually depends only on the homotopy type of $|K|$ because that is the case for the Betti numbers.

Theorem 7.2.3.10 suggests the following definition of a well-known invariant of a polyhedron.

**Definition.** If **X** is a polyhedron, the *Euler-Poincaré characteristic* of **X**, $\chi(\mathbf{X})$, is defined by

$$\chi(\mathbf{X}) = \sum_{q=0}^{\dim X} (-1)^q \beta_q(\mathbf{X}).$$

The typical way to compute the Euler-Poincaré characteristic of a polyhedron **X** is of course to use a simplicial complex K that triangulates **X** and use the numbers $n_q(K)$. This is also how it is often defined. Our definition has the advantage that the property of it being an intrinsic invariant of a polyhedron that is independent of any triangulation is built into the definition.

### 7.2.4 Cell Complexes

The homology theory we developed was based on simplices, but as we have mentioned before, we could have used other spaces as our basic building blocks, such as n-dimensional cubes, for example. The main advantage of simplices is a theoretical one. They simplify some formulas and constructions. A big practical disadvantage of simplices, however, is the fact that the simplicial complexes that triangulate spaces typically contain a great many simplices. Even a simplicial complex that triangulates a simple space such as the basic n-dimensional simplex already has an exponential number of simplices (as a function of n). Any algorithm for computing homology groups based on simplices would quickly be overwhelmed by their number for all but relatively low-dimensional spaces. Fortunately, one can define homology groups based on more efficient decompositions of spaces.

**Definition.** An *open k-cell* is any space **c** that is homeomorphic to $\mathbf{R}^k$. The integer k is called its *dimension* is denoted by dim **c**. An *open cell* is an open c-cell for some k. A *cell decomposition* of a space is a collection of disjoint open cells whose union is the space.

Note that the dimension of an open cell is well defined by Theorem 7.2.3.5. A straightforward generalization of simplicial complexes is to look for cell decompositions where we allow curved cells rather than just linear cells like the simplices. Actually, we shall go a step further.

**Definition.** A map $f:(\mathbf{X},\mathbf{A}) \to (\mathbf{Y},\mathbf{B})$ is called a *relative homeomorphism* if $f:\mathbf{X} \to \mathbf{Y}$ is a continuous map that maps $\mathbf{X} - \mathbf{A}$ homeomorphically onto $\mathbf{Y} - \mathbf{B}$.

Note that the relative homeomorphism f need not be a homeomorphism because it might not be one-to-one on **A**.

**Definition.** Let **A** be a closed subspace of **X**. We shall say that **X** is *obtained from **A** by adjoining k n-cells* $\mathbf{c}_i^n$, $n \geq 0$ and $0 \leq i < k$ (we allow $k = \infty$), if the following holds:

(1) Each $\mathbf{c}_i^n$ is a subspace of **X** and $\mathbf{X} = \mathbf{A} \cup \mathbf{c}_0^n \cup \mathbf{c}_1^n \cup \ldots$.
(2) If

$$\dot{\mathbf{c}}_i^n = \mathbf{c}_i^n \cap \mathbf{A},$$

then $\mathbf{c}_i^n - \dot{\mathbf{c}}_i^n$ and $\mathbf{c}_j^n - \dot{\mathbf{c}}_j^n$ are disjoint for $i \neq j$.
(3) **X** has the weak topology with respect to the sets **A** and $\mathbf{c}_i^n$.
(4) For each n-cell $\mathbf{c}_i^n$, there exists a relative homeomorphism

$$f_i : (\mathbf{D}^n, \mathbf{S}^{n-1}) \to (\mathbf{c}_i^n, \dot{\mathbf{c}}_i^n)$$

that maps $\mathbf{S}^{n-1}$ onto the set $\dot{\mathbf{c}}_i^n$. The map $f_i$ is called a *characteristic map* for the n-cell $\mathbf{c}_i^n$ and $g_i = f_i \,|\, S^{n-1}$ is called the *attaching map* for the n-cell $\mathbf{c}_i^n$.

Condition (4) justifies us calling $\mathbf{c}_i^n$ an *n-cell* or *cell* or *closed n-cell* (one can show that $\mathbf{c}_i^n$ is a closed subset of **X**). Note however that, although $\mathbf{c}_i^n - \dot{\mathbf{c}}_i^n$ is an open n-cell since it is homeomorphic to $\mathbf{R}^n$, $\mathbf{c}_i^n$ may not be homeomorphic to the closed disk $\mathbf{D}^n$ because f is not required to be one-to-one on $\mathbf{S}^{n-1}$.

Like in Section 5.3, one can think of these attaching maps as specifying a way to glue an n-disk $\mathbf{D}^n$ to a space along its boundary $\mathbf{S}^{n-1}$. This continues the cut-and-paste paradigm from the last chapter except that we are not doing any "cutting" right now. An alternate description of **X** is that

$$\mathbf{X} = \mathbf{A} \cup_{g_0} \mathbf{D}^n \cup_{g_1} \mathbf{D}^n \cup \ldots.$$

**7.2.4.1. Example.** The n-sphere $\mathbf{S}^n$ can be thought of as a space obtained from a point by attaching an n-cell using an attaching map that collapses the boundary of the n-cell to a point. For example, consider $\mathbf{S}^1$. A natural characteristic map is

$$f: \ \mathbf{D}^1 = [-1,1] \to \mathbf{S}^1, \ f(t) = (\cos t\pi, \sin t\pi),$$

which shows that $\mathbf{S}^1$ can be thought of as the point $(-1,0)$ with a 1-cell attached.

**Definition.** A *CW complex* C is a Hausdorff space **X** together with a sequence of closed subspaces $\mathbf{X}^n$ of **X**, $n = -1, 0, 1, \ldots$, satisfying

(1) $\phi = \mathbf{X}^{-1} \subset \mathbf{X}^0 \subset \mathbf{X}^1 \subset \ldots$

(2) $\mathbf{X} = \bigcup_{n=0}^{\infty} \mathbf{X}^n$.

(3) Each $\mathbf{X}^n$ is obtained from $\mathbf{X}^{n-1}$ by adjoining n-cells $\mathbf{c}_i^n$ via characteristic maps $f_{n,i}$. The n-cells $\mathbf{c}_i^n$ are called the (*closed*) *n-cells* of C and $\mathbf{c}_i^n - \dot{\mathbf{c}}_i^n = \mathbf{c}_i^n - \mathbf{X}^{n-1}$, the *open n-cells*.
(4) **X** has the weak topology with respect to the subspaces $\mathbf{X}^n$.

The space $\mathbf{X}$ is called the *underlying space* of the CW complex C and denoted by $|C|$. If the subspace structure on $\mathbf{X}$ defined by C is clear from the context, one often uses the phrase "the CW complex $\mathbf{X}$" to refer to C. We define $C^n$, called the *n-skeleton* of C, to be the CW complex with underlying space $\mathbf{X}^n$ and sequence of closed subspaces

$$\mathbf{X}^{-1} \subset \mathbf{X}^0 \subset \mathbf{X}^1 \subset \ldots \quad \mathbf{X}^{n-1} \subset \mathbf{X}^n \subset \mathbf{X}^n \subset \ldots.$$

The subspace $\mathbf{X}^n$ is usually also called the *n-skeleton* of C. A CW complex C is *finite* if it has only finitely many cells. It has *dimension* n if it has at least one n-cell but no m-cells for m > n (equivalently, $|C^n| = |C^{n+1}| = \ldots$). If C has dimension n for some n (denoted by dim C = n), then it is said to be *finite dimensional*, otherwise, it is said to be *infinite dimensional*. If the characteristic maps are all homeomorphisms, then the CW complex is said to be *regular*, otherwise, it is *irregular*.

CW complexes were first defined by J.H.C. Whitehead in 1949. The "C" stands for "closure finite" and the "W" for "weak topology." The term *closure finite* is the technical term for the property that every closed n-cell of a CW complex C meets only a finite number of open cells in $|C^{n-1}|$. For a proof that this property is satisfied by a CW complex as defined above see [LunW69]. Whitehead's definition of a CW complex actually differed slightly from the one given here. He started with abstract cell decompositions for which that property had to be stipulated. One can easily show that condition (4) of the definition is trivially satisfied for finite CW complexes and can be omitted if one restricts oneself to such complexes.

A CW complex defines a cell decomposition for a space. One normally thinks of a CW complex as a space where one start with some 0-cells (points), then attaches some 1-cells, then some 2-cells, and so on. Cells are to CW complexes what simplices are to simplicial complexes. (Note that the initial set of points can also be thought of in terms of having attached some 0-cells to an initially empty space.) Clearly, every simplicial complex is a regular CW complex because it defines an obvious sequence of cells and skeletons. The main difference between the cell decomposition induced on a space $\mathbf{X}$ by a regular CW complex and a triangulation of $\mathbf{X}$ is that cells have a flexible number of faces and are potentially "curved" from the start.

Figure 7.8 shows the inductive aspect of the definition of a regular CW complex by showing the steps that represent a disk as a regular CW complex. We started with two points and then attached two 1-cells and one 2-cell. It is easy to construct a regular cell decomposition for a torus by dividing a rectangle into four equal subrectangles and identifying the boundary pieces appropriately. This structure will have four 0-cells, eight 1-cells, and four 2-cells and is quite an improvement, in terms of numbers of cells, over the standard triangulation of the torus shown in Figure 7.5. In fact, we can even do better. Figure 7.9 shows a CW complex whose underlying space is the torus and which consists of four cells – one 0-cell $\mathbf{a}$, two 1-cells $\mathbf{b}$ and $\mathbf{c}$, and one 2-cell. The boundary of the 2-cell in Figure 7.9(a) is mapped onto the 1-skeleton $\mathbf{a} \cup \mathbf{b} \cup \mathbf{c}$ in Figure 7.9(b) by mapping the edges $\mathbf{b}_1$ and $\mathbf{c}_1$ to $\mathbf{b}$ and $\mathbf{c}$, respectively, using the orientation indicated by the arrows. This will send the vertices $\mathbf{a}$ and $\mathbf{a}_i$ to $\mathbf{a}$. Note that if we were to cut along the circles $\mathbf{b}$ and $\mathbf{c}$ in the torus we would unfold it to a rectangle.

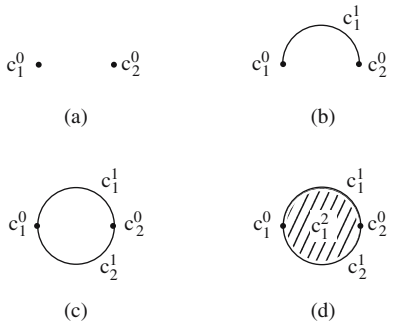



**Figure 7.8.** Generating a cell decomposition of a disk.

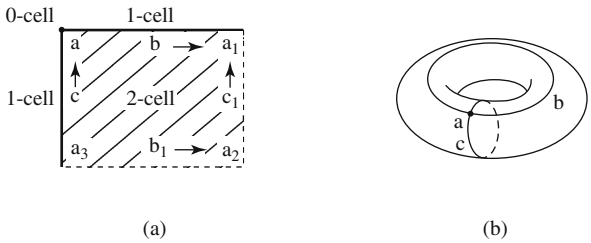



**Figure 7.9.** A compact cell decomposition for the torus.

**Definition.** A *subcomplex* of a CW complex C is a CW complex L such that

(1) $|L|$ is a closed subspace of $|C|$,
(2) $|L^n| = |L| \cap |C^n|$, and
(3) each cell of L is a cell of C.

**Definition.** A CW complex is *locally finite* if each of its closed cells meets only a finite number of other cells (of **any** dimension). It is *normal* if each closed cell is a subcomplex.

**Definition.** Let C and C′ be CW complexes. A continuous map $f : |C| \to |C'|$ is called a *cellular map* if $f(C^q) \subseteq (C'^q)$ for all q.

Here are some basic facts that hold for CW complexes. Some are easy to prove and are good exercises for the reader. The proofs of those that are not can be found in [LunW69] or [Jäni84].

(1) Polyhedra are CW complexes. Every finite CW complex has the homotopy type of a polyhedron and every finite regular normal CW complex can be triangulated. We can drop the finiteness conditions here if we allow infinite simplicial complexes. CW complexes are more general than polyhedra however. In particular, [LunW69] gives an example of a finite three-dimensional CW complex that cannot be triangulated.

(2) If C is a CW complex, then its skeletons $C^n$ are subcomplexes.
(3) Arbitrary unions and intersections of subcomplexes are subcomplexes.
(4) Every compact subset of a CW complex lies in a finite subcomplex.
(5) A CW complex is compact if and only if it is finite.
(6) If **X** is a locally finite CW complex, then any union of closed cells is a closed subset of **X**.
(7) If **X** is a CW complex and **Y** is an arbitrary topological space, then a map $f: \mathbf{X} \to \mathbf{Y}$ is continuous if and only if $f|\sigma$ is continuous for every closed cell $\sigma$ in **X**.
(8) Let **X** and **Y** be CW complexes and let **A** be a subcomplex of **X**. Any continuous map $f: \mathbf{X} \to \mathbf{Y}$ that is cellular on **A** is homotopic to a cellular map $g: \mathbf{X} \to \mathbf{Y}$ relative to **A**, that is, there is a homotopy $h: \mathbf{X} \times \mathbf{I} \to \mathbf{Y}$ so that $h(\mathbf{x},0) = f(\mathbf{x})$, $h(\mathbf{x},1) = g(\mathbf{x})$, and $h(\mathbf{a},t) = f(\mathbf{a})$, for all $\mathbf{a} \in \mathbf{A}$ and $t \in \mathbf{I}$.
(9) If the spaces **X** and **Y** are finite CW complexes, then so is $\mathbf{X} \times \mathbf{Y}$.
(10) If the subspace **A** is a subcomplex of a CW complex **X**, then the cells in $\mathbf{X} - \mathbf{A}$ and one 0-cell corresponding to **A** define a cell decomposition for $\mathbf{X}/\mathbf{A}$ that make it into a CW complex.
(11) Attaching cells to a CW complex produces a CW complex.
(12) Let **X** and **Y** be CW complexes and let **A** be a subcomplex of **X**. If $f: \mathbf{A} \to \mathbf{Y}$ is a cellular map, then $\mathbf{Y} \cup_f \mathbf{X}$ is a CW complex.
(13) A CW complex is a paracompact space and hence also a normal space.
(14) A connected CW complex is metrizable if and only if it is locally finite.

The cell decompositions of CW complexes are usually more natural decompositions of a space than triangulations and there are typically substantially fewer cells in a cell decomposition than simplices in a triangulation, but there are a number of other advantages to using CW complexes, making them the spaces of choice for topologists. For example, properties (9)–(11) are easy for CW complexes and would be more complicated for simplicial complexes because it would involve subdivisions.

It is possible to define a homology theory for a finite regular normal CW complex C (actually for any CW complex, but this gets more involved) merely by copying what was done in the case of simplicial complexes. For details see [CooF67]. In other words, one can define the notion of an oriented cell and then the group of q-chains, $C_q(C)$, of C is obtained by taking formal linear combinations of oriented q-cells in C. There is also a natural boundary map

$$\partial_q : C_q(C) \to C_{q-1}(C)$$

and well-defined group

$$H_q(C) = \frac{\ker \partial_q}{\operatorname{im} \partial_{q+1}}$$

called the qth homology group of C.

Using the fact that any continuous map $f: |C| \to |C'|$ between the underlying spaces of two CW complexes C and C′ is homotopic to a cellular map, there is an induced natural homomorphism

$$f_{*q} : H_q(C) \to H_q(C').$$

The groups $H_q(C)$ and homomorphisms $f_{*q}$ satisfy all the properties that their simplicial analogs did. Finally, one can prove the important theorem that asserts that if C is a finite regular normal CW complex and K is a simplicial complex with $|C|$ homeomorphic to $|K|$, then $H_q(C)$ is isomorphic to $H_q(K)$. This means that one can obtain the homology groups of a polyhedron either from a simplicial or a cell structure.

**Definition.** If C is a finite CW complex, then let $n_q(C)$ denote the number of q-cells in C and define the *Euler-Poincaré characteristic* of C, $\chi(C)$, by

$$\chi(C) = \sum_{q=0}^{\dim C} (-1)^q n_q(C).$$

The Euler-Poincaré characteristic of a CW complex is really nothing new.

**7.2.4.2. Theorem.** If C is a finite regular normal CW complex, then $\chi(C) = \chi(|C|)$.

**Proof.** See [CooF67]. Note that the hypothesis implies that $|C|$ is a polyhedron.

We know that the Euler-Poincaré characteristic and the dimension of a polyhedron is a topological invariant. There is an interesting related fact.

**Definition.** If $f(x_0,x_1,\ldots)$ is any function of the indeterminates $x_i$ and if C is a CW complex, then define

$$f(C) = f(n_0(C), n_1(C), \ldots).$$

We shall say that f is *topologically invariant function* if $f(C) = f(C')$ for all CW complexes C and C′ with homeomorphic underlying spaces.

**7.2.4.3. Example.** If

$$f(x_0, x_1, \ldots) = x_0 - x_1x_3 + x_7^2 \quad \text{and} \quad g(x_0, x_1, \ldots) = \sum_{q=0}^{\infty} (-1)^q x_q,$$

then

$$f(C) = n_0(C) - n_1(C)n_3(C) + n_7(C)^2 \quad \text{and} \quad g(C) = \chi(C).$$

**7.2.4.4. Theorem.** The only topologically invariant functions $f(x_0,x_1,\ldots)$ on CW complexes are those that are functions of the Euler-Poincaré characteristic and the dimension, that is, if C and C′ are CW complexes with dim C = dim C′ and $\chi(C) = \chi(C')$, then $f(C) = f(C')$.

**Proof.** See [AgoM76].

This leads to the next topic in this section. Given a polyhedron **X**, what is the fewest number of cells needed in a cellular decomposition of **X**? The answer to this question is not only important for computation purposes but also to the topological classification of polyhedra. Exercise 6.5.5 gave a partial answer in some special cases to the corresponding question for simplicial complexes.

**Definition.** Let $\mathbf{c}^n$ be an n-cell of a CW complex C that does not belong to any higher-dimensional cell of C. If $\mathbf{c}^{n-1}$ is an (n – 1)-cell of C that is contained in $\mathbf{c}^n$ but to no other higher-dimensional cell of C, then $\mathbf{c}^{n-1}$ is called a *free* cell of C.

Let C be a CW complex. Let $\mathbf{c}^n$ be a top-dimensional cell in C that contains a free (n – 1)-dimensional cell $\mathbf{c}^{n-1}$. Let C′ be the subcomplex one obtains after removing the cells $\mathbf{c}^n$ and $\mathbf{c}^{n-1}$, that is,

$$|C'| = |C| - \left(\mathbf{c}^n - \dot{\mathbf{c}}^n\right) - \left(\mathbf{c}^{n-1} - \dot{\mathbf{c}}^{n-1}\right).$$

**Definition.** We shall say that C′ is obtained from C by an *elementary collapse* from $\mathbf{c}^{n-1}$ through $\mathbf{c}^n$. Conversely, we say that C is obtained from C′ by an *elementary expansion* using the cell pair $(\mathbf{c}^n, \mathbf{c}^{n-1})$.

See Figure 7.10. An elementary collapse through a free cell gives rise to a natural attaching maps, so that C can be thought of as being obtained from C′ by attaching an n-cell.

**Definition.** Let C and C′ be CW complexes. We say that C *collapses* to C′ and write $C \downarrow C'$ if there exists a sequence of CW complexes $C_0 = C \supset C_1 \supset \ldots \supset C_n = C'$ so that $C_{i+1}$ is obtained from $C_i$ via an elementary collapse. In that case we also say that C′ *expands* to C.

Figure 7.10 shows a sequence of elementary collapses that collapse a disk to a point. Figure 7.11 shows how a cell decomposition of an annulus can be collapsed to a circle. The numbers in the 2-cells of Figure 7.11(a) indicate the order of their collapse, which are then to be followed by the 1-cell collapses whose order is indicated by the numbers in Figure 7.11(b). We end up with the circle in Figure 7.11(c).

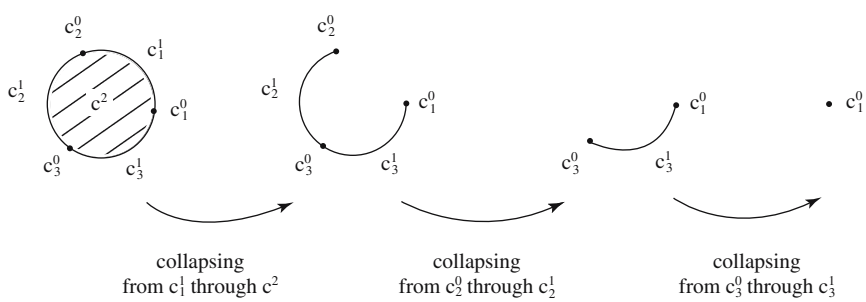



**Figure 7.10.** Collapsing a disk to a point.

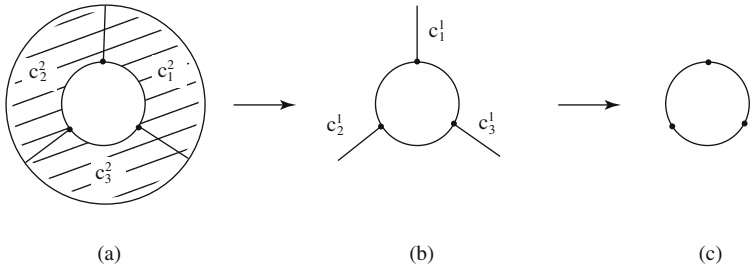



**Figure 7.11.** Collapsing an annulus to a circle.

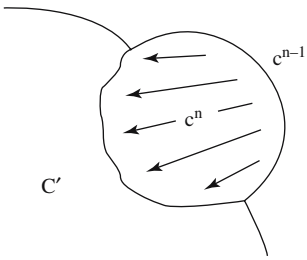



**Figure 7.12.** Retracting from free cell $\mathbf{c}^{n-1}$ through cell $\mathbf{c}^n$.

**7.2.4.5. Theorem.** Let C and C′ be CW complexes. If C collapses to C′, then $|C'|$ is a deformation retract of $|C|$.

**Proof.** It suffices to prove the theorem in the case where C′ is obtained from C by an elementary collapse from an (n − 1)-cell $\mathbf{c}^{n-1}$ through an n-cell $\mathbf{c}^n$ because the general case can be proved by induction. See Figure 7.12. The result follows from the fact that it is easy to construct a deformation retraction of $\mathbf{D}^n$ to $\mathbf{S}_-^{n-1}$ by "pushing down" through $\mathbf{D}^n$ from $\mathbf{S}_+^{n-1}$ and is left as an exercise.

**7.2.4.6. Corollary.** Let C and C′ be CW complexes. If C collapses to C′, then C and C′ have the same homotopy type. In fact, the inclusion map of $|C'|$ in $|C|$ is a homotopy equivalence.

**Proof.** This follows from Theorems 7.2.4.5 and 5.7.7.

Corollary 7.2.4.6 just reinforces what we have said before, namely, that homology and homotopy invariants are not good enough for detecting when spaces are homeomorphic.

Returning to the problem of finding a minimal cell decomposition of a polyhedron $\mathbf{X}$, the idea is to start with any CW complex C with $|C| = \mathbf{X}$ and then

(1) Collapse C as much as possible to a CW complex $C_1$.
(2) Pick a cell $\mathbf{c}_1$ in $C_1$ that does not belong to any other cell in $C_1$, remove it from $C_1$, and collapse the remainder as much as possible to another CW complex $C_2$.
(3) Repeat step (2) as long as there are cells $\mathbf{c}_i$ to pick.

**Table 7.2.4.1** Minimal cell decompositions.

| X | $n_0(\mathbf{X})$ | $n_1(\mathbf{X})$ | $n_2(\mathbf{X})$ |
|---|---|---|---|
| Orientable surface (genus k) | 1 | 2k | 1 |
| Nonorientable surface (genus k) | 1 | k | 1 |

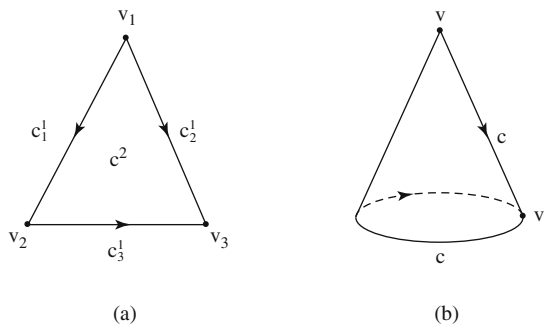



**Figure 7.13.** The dunce hat.

This will lead to a cell complex built up from a sequence of cells $\mathbf{c}_1'$, $\mathbf{c}_2'$, . . . , $\mathbf{c}_k'$ derived from the $\mathbf{c}_i$ whose underlying space has the same homotopy type as $|C|$. This approach leads to Table 7.2.4.1 that lists the number of cells in a minimal cell decompositions for compact connected surfaces (without boundary).

Will the construction above lead to a minimal cell decomposition? In low dimensions, the answer is yes, but in general, the answer is no. To see why this is so, consider Figure 7.13 which shows part of the steps in the construction of the space called the *dunce hat*. Start with the triangle shown in Figure 7.13(a) and identify the three edges $\mathbf{c}_1^1$, $\mathbf{c}_2^1$, and $\mathbf{c}_3^1$ using the orientation of the edges shown by the arrows. Figure 7.13(b) does not yet show the final picture because one still needs to identify the two edges marked **c**. The dunce hat is clearly not collapsible because it has no free edges but one can prove that it is contractible. A consequence is that there are cellular decompositions so that if one is not careful about choosing the sequence of collapses, one might not end up with a minimal cell decomposition. This can happen even in the case of a simple space such as $\mathbf{D}^n$. There are cellular decompositions of $\mathbf{D}^n$, $n > 2$, with the property that **no** sequence of collapses will end up with a point. See [BurM71].

If we look back over the examples we have given of topological spaces since we started talking about topology in Chapter 5, we can see that, as far as manifolds were concerned, they have been pretty limited and were restricted to such "standard" spaces as surfaces, $\mathbf{R}^n$, $\mathbf{S}^n$, $\mathbf{D}^n$, $\mathbf{P}^n$, etc. We got the most variety of spaces from surfaces, but our classification theorems showed that they also fit into simple patterns. Once one gets above dimension 2, however, things change drastically and a neat classification is no longer possible. Just so that the reader does not think that there is nothing new out there and gets bored with all the current examples, we finish this section by defining a well-known class of three-dimensional manifolds that are quite different from the ones we have seen up to now. Fortunately, they are relatively easy to describe in terms of a cell structure.

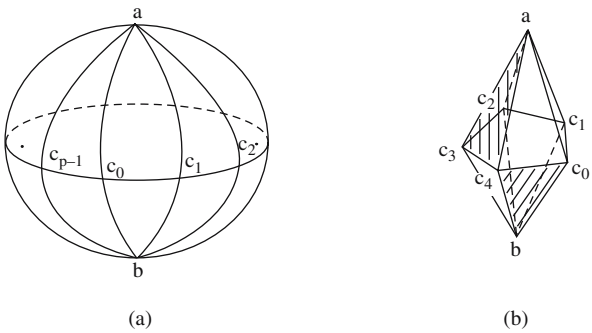



**Figure 7.14.** The lens paces L(p,q) and L(5,1).

**Definition.** Let p and q be relatively prime positive integers and assume that $0 \le q \le p/2$. Define the *lens space* L(p,q) as follows: Let S be the reflection in $\mathbf{R}^3$ about the x-y plane and let R be the rotation about the z-axis through an angle of $2\pi q/p$ radians. Then

$$L(p,q) = \mathbf{D}^3/\sim,$$

where ~ is the equivalence relation induced by the identification of $\mathbf{x} \in \mathbf{S}^2_+$ with $R(S(\mathbf{x}))$. In other words, we are identifying the point $\mathbf{x}$ in the upper hemisphere with the point in the lower hemisphere obtained by reflecting $\mathbf{x}$ about the xy-plane and then rotating by $2\pi q/p$. We are not identifying any points in the interior of the disk $\mathbf{D}^3$.

Here is another description of L(p,q). See Figure 7.14(a). Let

$$\mathbf{c}_k = (\cos 2\pi k/p, \sin 2\pi k/p, 0), \quad k = 0,1,\ldots,p-1.$$

The points $\mathbf{c}_k$ divide the equator of the sphere into p equal arcs. Let $\mathbf{a} = \mathbf{e}_3$ and $\mathbf{b} = -\mathbf{e}_3$ be the north and south pole of the sphere, respectively. We get the following cell decomposition of $\mathbf{D}^3$:

0-cells: $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}_k$
1-cells: the great arcs from $\mathbf{a}$ to $\mathbf{c}_k$ and from $\mathbf{b}$ to $\mathbf{c}_k$ and the arcs along the unit circle from $\mathbf{c}_k$ to $\mathbf{c}_{k+1}$
2-cells: 2p curved triangles, denoted by $\mathbf{v}\mathbf{c}_k\mathbf{c}_{k+1}$, bounded by the arcs from $\mathbf{v}$ to $\mathbf{c}_k$, from $\mathbf{c}_k$ to $\mathbf{c}_{k+1}$, and from $\mathbf{c}_{k+1}$ back to $\mathbf{v}$, where $\mathbf{v}$ is $\mathbf{a}$ or $\mathbf{b}$
3-cells: $\mathbf{D}^3$

(All indices are taken modulo p.) L(p,q) is now the disk $\mathbf{D}^3$ where we identify the curved triangle $\mathbf{a}\mathbf{c}_k\mathbf{c}_{k+1}$ in the upper hemisphere with the curved triangle in the lower hemisphere $\mathbf{b}\mathbf{c}_{k+q}\mathbf{c}_{k+q+1}$ and the vertices are identified in the order listed. Figure 7.14(b) shows the case of L(5,1) where we have linearized the construction and have replaced the sphere by a suspension of a five-sided polygon. The shaded triangle $\mathbf{a}\mathbf{c}_3\mathbf{c}_4$ gets identified with the shaded triangle $\mathbf{b}\mathbf{c}_4\mathbf{c}_0$.

Note that L(1,q) is the sphere $\mathbf{S}^3$, but this case has to be treated in a slightly special way if one wants a nice cell decomposition. One has to add another vertex at $-\mathbf{e}_1$.

Lens spaces were first defined by Tietze ([Tiet08]). We list some of their properties below:

(1) They are closed compact three-dimensional manifolds.
(2) They can be triangulated.
(3) L(2,1) is homeomorphic to $\mathbf{P}^3$.
(4) The homology groups of L(p,q) are

$$H_i(L(p,q)) \approx \begin{Bmatrix} \mathbf{Z} & i=0 \\ \mathbf{Z}_p & i=1 \\ 0 & i=2 \\ \mathbf{Z} & i=3 \end{Bmatrix}.$$

(5) $\chi(L(p,q)) = 0$.
(6) L(p,q) is homeomorphic to L(p,q′) if and only if

$$q' \equiv \pm q^{\pm 1} (\text{mod } p).$$

(7) L(p,q) and L(p,q′) have the same homotopy type if and only if qq′ or –qq′ is a quadratic residue modulo p.

Properties (1)–(3) are easy to prove. Figure 7.14(b) should make clear what one needs to do for (2). For properties (4) and (5) use the cell decomposition of L(p,q) induced by the cell decomposition of $\mathbf{D}^3$ described above to compute its homology groups. Property (6) was proved by Reidemeister ([Reid35]). Property (7) was proved by Whitehead ([Whit41]). More details about lens spaces can also be found in [SeiT80] and [HilW60].

Although it is obvious that homotopy equivalence is a weaker relation than homeomorphism (consider a disk and a point or, more generally, any deformation retract), it is not so obvious with respect to some types of spaces like manifolds without boundary. This makes the next example and lens spaces all the more interesting.

**7.2.4.7. Example.** A consequence of properties (6) and (7) is that the 3-manifolds L(7,1) and L(7,2) have the same homotopy type but are not homeomorphic.

## 7.2.5 Incidence Matrices

This section discusses the so-called incidence matrices. These matrices played an important role in the history of combinatorial topology. An excellent detailed account of these matrices can be found in [Cair68]. Computers can easily use them to compute homology groups.

Let K be a simplicial complex of dimension n and, as before, let $n_q = n_q(K)$ be the number of q-simplices in K. For each q, assume that we have chosen an orientation for all the q-simplices in K and let

$$S_q^+ = \left\{\left[\boldsymbol{\sigma}_1^q\right], \left[\boldsymbol{\sigma}_2^q\right], \ldots, \left[\boldsymbol{\sigma}_{n_q}^q\right]\right\}$$

be the set of these oriented q-simplices of K. The set $S_q^+$ will be a basis for the free abelian group $C_q(K)$. Define integers $\varepsilon_{ij}^q$ by the equation

$$\partial_{q+1}\left[\boldsymbol{\sigma}_j^{q+1}\right] = \sum_{i=1}^{n_q} \varepsilon_{ij}^q \left[\boldsymbol{\sigma}_i^q\right]$$

and note that

$$\varepsilon_{ij}^q = \begin{Bmatrix} \pm 1 & \text{if } \boldsymbol{\sigma}_i^q \text{ is a face of } \boldsymbol{\sigma}_{j+1}^q \\ 0 & \text{otherwise} \end{Bmatrix}.$$

**Definition.** The integer $\varepsilon_{ij}^q$ is called the *incidence number* of $[\boldsymbol{\sigma}_i^q]$ and $[\boldsymbol{\sigma}_{j+1}^q]$. The *qth incidence matrix* $E^q$, $0 \le q < \dim K$, of K is defined to be the $(n_q \times n_{q+1})$-matrix

$$E^q = \left(\varepsilon_{ij}^q\right) = \begin{array}{c} \\ \left[\boldsymbol{\sigma}_1^q\right] \\ \left[\boldsymbol{\sigma}_2^q\right] \\ \vdots \\ \left[\boldsymbol{\sigma}_{n_q}^q\right] \end{array} \begin{array}{c} \begin{array}{cccc} \left[\boldsymbol{\sigma}_1^{q+1}\right] & \left[\boldsymbol{\sigma}_2^{q+1}\right] & \cdots & \left[\boldsymbol{\sigma}_{n_{q+1}}^{q+1}\right] \end{array} \\ \begin{pmatrix} \varepsilon_{11}^q & \varepsilon_{12}^q & \cdots & \varepsilon_{1n_{q+1}}^q \\ \varepsilon_{21}^q & \varepsilon_{22}^q & \cdots & \varepsilon_{2n_{q+1}}^q \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n_q 1}^q & \varepsilon_{n_q 2}^q & \cdots & \varepsilon_{n_q n_{q+1}}^q \end{pmatrix} \end{array}$$

whose rows and columns are indexed by the elements of $S_q^+$ and $S_{q+1}^+$, respectively.

**7.2.5.1. Example.** Suppose that $K = \langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2 \rangle$ and that we have chosen the $S_q^+$ as follows:

$$\begin{aligned} S_0^+ &= \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2\}, \\ S_1^+ &= \{[\mathbf{v}_0\mathbf{v}_1], [\mathbf{v}_1\mathbf{v}_2], [\mathbf{v}_0\mathbf{v}_2]\}, \\ S_2^+ &= \{[\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2]\}. \end{aligned}$$

The incidence matrices of K are then given by

$$E^1 = \begin{array}{c} \\ [\mathbf{v}_0\mathbf{v}_1] \\ [\mathbf{v}_1\mathbf{v}_2] \\ [\mathbf{v}_0\mathbf{v}_2] \end{array} \begin{array}{c} [\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2] \\ \begin{pmatrix} +1 \\ +1 \\ -1 \end{pmatrix} \end{array} \quad \text{and} \quad E^0 = \begin{array}{c} \\ \mathbf{v}_0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \end{array} \begin{array}{c} \begin{array}{ccc} [\mathbf{v}_0\mathbf{v}_1] & [\mathbf{v}_1\mathbf{v}_2] & [\mathbf{v}_0\mathbf{v}_2] \end{array} \\ \begin{pmatrix} +1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & +1 & +1 \end{pmatrix} \end{array}.$$

Next, we define the incidence matrices with respect to arbitrary bases

$$\{c^q\} = \{c_1^q, c_2^q, \ldots, c_{n_q}^q\}$$

for the free abelian groups $C_q(K)$. Since we have bases, there are unique integers $\eta_{ij}^q$ such that

$$\partial_{q+1}\left(c_j^{q+1}\right) = \sum_{i=1}^{n_q} \eta_{ij}^q c_j^q.$$

**Definition.** The *qth incidence matrix* of K *with respect to the bases* $\{c^{q+1}\}$ and $\{c^q\}$ is defined to be the $(n_q \times n_{q+1})$-matrix

$$\left(g_{ij}^q\right) = \begin{array}{c} \\ \vdots \\ c_i^q \\ \vdots \end{array} \begin{array}{c} \cdots \quad c_j^{q+1} \quad \cdots \\ \begin{pmatrix} \vdots & \\ \cdots & \eta_{ij}^q & \cdots \\ \vdots & \end{pmatrix} \end{array}.$$

Let c be an arbitrary (q + 1)-chain. If we express c with respect to the basis $\{c^{q+1}\}$, then

$$c = \sum_{j=1}^{n_{q+1}} a_j c_j^{q+1}$$

for some unique integers $a_j$ and

$$\partial_{q+1}(c) = \sum_{j=1}^{n_{q+1}} a_j \partial_{q+1}\left(c_j^{q+1}\right) = \sum_{i=1}^{n_q}\left(\sum_{j=1}^{n_{q+1}} a_j \eta_{ij}^q\right) c_i^q.$$

This shows that the boundary homomorphisms $\partial_q$ of K, and hence the homology groups of K, are completely determined once the incidence matrices are known with respect to some bases. Our goal in the remainder of this section is to show how the homology groups of K can be computed from knowledge of the basic incidence matrices $E^q$ of K alone.

**7.2.5.2. Lemma.** Choose a basis for each group $C_q(K)$. The matrix product of any two successive incidence matrices with respect to any such choice of bases is the zero matrix. Using the notation above, this means that for all q

$$\left(\eta_{ij}^{q-1}\right)\left(\eta_{ij}^q\right) = (n_{q-1} \times n_{q+1})\text{-matrix of zeros.}$$

**Proof.** This is a straightforward consequence of Lemma 7.2.1.3(2).

Lemma 7.2.5.2 and some purely algebraic manipulations of matrices lead to

**7.2.5.3. Theorem.** It is possible to choose bases for all the groups $C_q(K)$ simultaneously with respect to which the qth incidence matrix has the normalized form

$$N_*^q = \begin{array}{c} \begin{array}{cc} \overbrace{\qquad\qquad}^{n_{q+1}-\gamma_q \text{ columns}} & \overbrace{\qquad\qquad\qquad\qquad}^{\gamma_q \text{ columns}} \end{array} \\ \left(\begin{array}{cccc} 0 & 0 & \delta_1^q & 0 \\ & & \delta_2^q & \\ & & & \ddots \\ 0 & 0 & 0 & \delta_{\gamma_q}^q \\ 0 & 0 & 0 & 0 \\ & & & \\ 0 & 0 & 0 & 0 \end{array}\right) \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} \gamma_q \text{ rows} \\ \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} n_q - \gamma_q \text{ rows} \end{array} \tag{7.3}$$

where

(1) the $\delta^q$'s are positive integers,
(2) $\delta_{i+1}^q$ divides $\delta_i^q$, and
(3) $n_q - \gamma_q \geq \gamma_{q-1}$.

**Outline of Proof.** Choose a basis

$$\{c^q\} = \{c_1^q, c_2^q, \ldots, c_{n_q}^q\}$$

for each group $C_q(K)$ and assume that $N^q$ is the qth incidence matrix with respect to these bases. We shall transform the matrices $N^q$ into the form (7.3) by appropriate changes to the bases and so we need to know how changes to bases affect the matrices. First of all, note that changing the basis $\{c^q\}$ clearly affects both $N^q$ and $N^{q-1}$.

**Claim 1.** The matrices $N^q$ have the following properties:

(a) Replacing

$$\{c_1^q, \ldots, c_i^q, \ldots, c_{n_q}^q\} \quad \text{by} \quad \{c_1^q, \ldots, -c_i^q, \ldots, c_{n_q}^q\}$$

corresponds to changing all signs in the ith column of $N^{q-1}$ and ith row of $N^q$.

(b) Replacing

$$\{c_1^q, \ldots, c_i^q, \ldots, c_j^q, \ldots, c_{n_q}^q\} \quad \text{by} \quad \{c_1^q, \ldots, c_j^q, \ldots, c_i^q, \ldots, c_{n_q}^q\}$$

corresponds to interchanging the ith and jth columns in $N^{q-1}$ and the ith and jth rows in $N^q$

(c) Replacing

$$\left\{c_1^q, \ldots, c_i^q, \ldots, c_{n_q}^q\right\} \quad \text{by} \quad \left\{c_1^q, \ldots, c_i^q + kc_j^q, \ldots, c_{n_q}^q\right\}$$

for some integers k and $i \neq j$ corresponds to replacing the ith column of $N^{q-1}$ by the ith column plus k times the jth column and replacing the jth row of $N^q$ by the jth row minus k times the ith row.

The proof of parts (a) and (b) of Claim 1 is easy and left to the reader. Part (c) is readily deduced from the identities

$$\partial_q\left(c_i^q + kc_j^q\right) = \partial_q\left(c_i^q\right) + k\partial_q\left(c_j^q\right)$$

and

$$\partial_{q+1}\left(c_s^{q+1}\right) = \sum_{t=1}^{n_q} \eta_{ts}^q c_t^q = \eta_{is}^q\left(c_i^q + kc_j^q\right) + \left(\eta_{js}^q - k\eta_{is}^q\right)c_j^q + \sum_{t=1, t\neq i,j}^{n_q} \eta_{ts}^q c_t^q.$$

Claim 1 is proved.

Now consider the 0th incidence matrix $E^0 = (\varepsilon_{ij}^0)$ of K. Note how each column has only two nonzero entries, namely, one +1 and one −1.

**Claim 2.** Any integer matrix $P = (p_{ij})$ having the property that each nonzero column has precisely two nonzero entries, one of which is +1 and the other is −1, can be transformed into a normalized matrix as defined by Theorem 7.2.5.3 via a sequence of matrix operations of the type described in Claim 1 above.

Claim 2 is proved by induction on the number of rows (or columns) in P. First, by interchanging rows, we may assume that $p_{11} = +1$. The next step is to zero out the first row past the first entry by a sequence of column operations. Specifically, if $p_{1j} = \pm 1$ for some $j > 1$, then replace the jth column of P by (jth column − $p_{1j}$ (1st column)). The new matrix $P' = (p'_{ij})$ will have $p'_{1j} = 0$ and either the jth column is zero or there are again just two nonzero entries, one +1 and the other −1. By a sequence of such operations we arrive at a matrix $P'' = (p''_{ij})$ such that $p''_{11} = 1$, $p''_{1j} = 0$, for $j > 1$, and $P''$ also satisfies the same hypotheses as the original matrix P. Now $p''_{i1}$ will equal −1 for some $i > 1$. Subtracting the 1st row from the ith row of $P''$, will give us a matrix $Q = (q_{ij})$, such that $q_{i1} = q_{1j} = 0$ for $i, j > 1$. The inductive hypothesis would then apply to the matrix $(q_{ij})_{2\leq i, 2\leq j}$ and finish the proof of Claim 2.

In order to prove our theorem we shall prove the following assertions for $k > 0$:

**Assertion $A_k$:** It is possible to choose bases for the groups $C_q(K)$ so that with respect to these bases the incidence matrices are

$$N_*^0, \ldots, N_*^{k-1}, N^k, E^{k+1}, \ldots, E^{n-1}.$$

(All other incidence matrices are uninteresting since they are zero.)

Note the following important property of the matrices $N^k$ in assertion $A_k$ that follows easily from the fact that, by Lemma 7.2.5.2, the product $N_*^{k-1}N^k$ is the zero matrix.

**Claim 3.** The last $\gamma_{k-1}$ rows of the matrix $N^k$ in assertion $A_k$ are zero.

We shall use induction on k to prove the assertions $A_k$ above. Assertion $A_1$ follows from Claims 1 and 2. These claims show that $E^0$ can be transformed into the normalized form $N_*^0$ by changing the basis of $C_0(K)$ and $C_1(K)$ appropriately. Although the 1st incidence matrix may have changed, the ith incidence matrices for $i \geq 2$ have not. One can check that the proof actually shows that $\delta_i^0 = 1$, for $1 \leq i \leq \gamma_0$.

Assume inductively that assertion $A_k$ is true for some $k > 0$. One can show that $N^k$ can be transformed into a normalized form such as is required for Theorem 7.2.5.3 by a sequence of matrix operations of the type described in Claim 1. This fact is a special case of a normalization theorem for matrices that is not hard but too long to reproduce here. A proof can, for example, be found in [Cair68].

Using Claim 3 we may assume that only the first $n_k - \gamma_{k-1}$ rows of $N^k$ will be affected as we transform the matrix to its normalized form. Translating the changes we make in the matrix into the corresponding changes in the basis for $C_k(K)$ and their effect on $N_*^{k-1}$, we can easily see that only the first $n_k - \gamma_{k-1}$ columns of $N_*^{k-1}$ are manipulated. Since these consist entirely of zeros, the matrix $N_*^{k-1}$ is left unchanged. Of course, the matrix $E^{k+1}$ will certainly have changed, but we have established assertion $A_{k+1}$. By induction, assertion $A_k$ is true for all $k > 0$. Assertion $A_n$ proves Theorem 7.2.5.3.

**A detailed version of the qth incidence matrix (7.3) of Theorem 7.2.5.3:** Recall that the rows and columns of the incidence matrices are indexed by the chains in the chosen basis of the appropriate chain groups of K. Therefore, assume that $0 \leq q \leq n$ and that the basis elements of $C_q(K)$ corresponding to the rows of $N_*^q$ have been labeled as follows:

The first $\gamma_q$ basis elements are labeled as $A_i^q$'s (note that $\gamma_{-1} = \gamma_n = 0$),
the last $\gamma_{q-1}$ are labeled as $C_i^q$'s,
the remaining $\beta = n_q - \gamma_q - \gamma_{q-1}$ basis elements, if there are any, are labeled as $B_i^q$'s, and
if the $\delta_i^q$ are the elements shown in the normalized matrix (7.3), then the integer $\rho_q$ is defined by

$$\rho_q = \max\left(0, \left\{i \mid \delta_i^q > 1\right\}\right).$$

With this notation, matrix (7.3) can be rewritten as

$$N_*^q = \begin{array}{c} \\ A_1^q \\ \vdots \\ A_{p_q}^q \\ A_{\rho_q+1}^q \\ \vdots \\ A_{\gamma_q}^q \\ B_1^q \\ \vdots \\ B_{\beta_q}^q \\ C_1^q \\ \vdots \\ C_{\gamma_{q-1}}^q \end{array} \begin{array}{c} \begin{array}{cccccccccccc} A_1^{q+1} & \cdots & A_{\gamma_{q+1}}^{q+1} & B_1^{q+1} & \cdots & B_{\beta_{q+1}}^{q+1} & C_1^{q+1} & \cdots & C_{\rho_q}^{q+1} & C_{\rho_q+1}^{q+1} & \cdots & C_{\gamma_q}^{q+1} \end{array} \\ \begin{pmatrix} & & & & & & \delta_1^q & & 0 & & & \\ & 0 & & & 0 & & & \ddots & & & 0 & \\ & & & & & & 0 & & \delta_{\rho_q}^q & & & \\ & & & & & & & & & 1 & & 0 \\ & 0 & & & 0 & & & 0 & & & \ddots & \\ & & & & & & & & & 0 & & 1 \\ & & & & & & & & & & & \\ & 0 & & & 0 & & & 0 & & & 0 & \\ & & & & & & & & & & & \\ & 0 & & & 0 & & & 0 & & & 0 & \\ & & & & & & & & & & & \end{pmatrix} \end{array}$$

The detailed version of the incidence matrices makes computing the homology groups of K easy because it tells us all we need to know about the groups $C_q(K)$ and homomorphisms $\partial_q$. In particular, it is easy to see that

$$\left\{\left[A_1^q\right], \ldots, \left[A_{\rho_q}^q\right], \left[B_1^q\right], \ldots, \left[B_{\beta_q}^q\right]\right\}$$

is a basis for $H_q(K)$. Also,

$o([A_i^q]) = \delta_i^q$, with the $\delta_i^q$'s being the torsion coefficients of $H_q(K)$,
$o([B_i^q]) = \infty$, and
$\operatorname{rk} H_q(K) = \beta_q$.

Finally, incidence matrices can be defined for CW complexes using their cells. All the information about homology groups that one could deduce in the simplicial case remain valid. This greatly simplifies computations because the dimensions of these matrices will be much smaller.

### 7.2.6 The Mod 2 Homology Groups

A more precise name for the homology groups of a simplicial complex K as defined in Section 7.2.1 is to call them the homology groups "with coefficients in **Z**." The reason is that chain group $C_q(K)$ consisted of formal **integer** linear combinations of oriented q-simplices. It is easy to generalize this.

Let G be an arbitrary abelian group and let $S_q$ again denote the set of oriented q-simplices of K. If $0 \le q \le \dim K$, define the group $C_q(K;G)$ of *q-chains with coefficients in G* by

$$C_q(K;G) = \{f : S_q \to G \mid \text{if } q \geq 1, \text{ then } f(-\alpha) = -f(\alpha) \text{ for all } \alpha \text{ in } S_q\}.$$

Clearly, $C_q(K)$ is just $C_q(K;\mathbf{Z})$. The elements of $C_q(K;G)$ can be thought of as formal sums

$$\sum_{\alpha \in S_q^+} g_\alpha \alpha, \tag{7.4}$$

where $g_\alpha \in G$, motivating the terminology. Elements are added by collecting the coefficients of the $\alpha$ and then adding them using the group addition. Pretty much everything we did earlier carries over without any problem if we simply replace $\mathbf{Z}$ by G. There is a boundary homomorphism

$$\partial_q : C_q(K;G) \to C_{q-1}(K;G)$$

and we can define the subgroup $Z_q(K;G)$ of *q-cycles* and the subgroup $B_q(K;G)$ of *q-boundaries with coefficients in G*. Again, $B_q(K;G) \subset Z_q(K;G)$. Finally,

**Definition.** The group

$$H_p(K;G) = \frac{Z_q(K;G)}{B_q(K;G)}.$$

is called the *qth homology group of K with coefficients in G*.

Simplicial maps $f : K \to L$ induce homomorphisms

$$f_{*q} : H_q(K;G) \to H_q(L;G),$$

and one can again show that the homology groups $H_q(K;G)$ are topological invariants and so we can associate a unique (up to isomorphism) group $H_q(\mathbf{X};G)$ to every polyhedron $\mathbf{X}$. They are important new invariants associated to a space even though it turns out that each is completely determined by $H_q(\mathbf{X};\mathbf{Z})$ and $H_{q-1}(\mathbf{X};\mathbf{Z})$ by the so-called "universal coefficient theorem." A very important special case is the case where $G = \mathbf{Z}_2$ and it is worthwhile to look at that case more carefully.

Let K be simplicial complex. Since +1 is the same as −1 in $\mathbf{Z}_2$, there is no need to orient the simplices of K to define $C_q(K;\mathbf{Z}_2)$. It follows that if $T_q$ is the set of all q-simplices in K, then $C_q(K;\mathbf{Z}_2)$ can be identified with the set of all maps

$$g : T_q \to \mathbf{Z}_2,$$

that is, an element of $C_q(K;\mathbf{Z}_2)$ (usually called a *mod 2 q-chain*) can be thought of as consisting of a linear sum of q-simplices of K. To add mod 2 q-chains we simply collect the coefficients of like simplices but must remember that $\boldsymbol{\sigma} + \boldsymbol{\sigma} = 0$ for every $\boldsymbol{\sigma} \in T_q$ because $2 = 0$ in $\mathbf{Z}_2$. For example, if K is the simplicial complex determined by the simplex $\mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3$, then

$$\mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_0, \mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_3 + \mathbf{v}_3\mathbf{v}_1 \in C_1(K;\mathbf{Z}_2)$$

and

$$(\mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_0) + (\mathbf{v}_1\mathbf{v}_2 + \mathbf{v}_2\mathbf{v}_3 + \mathbf{v}_3\mathbf{v}_1) = \mathbf{v}_0\mathbf{v}_1 + \mathbf{v}_3\mathbf{v}_1 + \mathbf{v}_2\mathbf{v}_3 + \mathbf{v}_2\mathbf{v}_0. \tag{7.5}$$

Recall our intuitive discussion of "holes" in Section 7.2.1. It is what we are doing now that makes precise the first approach to homology groups in that section.

The boundary map

$$\partial_q : C_q(K;\mathbf{Z}_2) \to C_{q-1}(K;\mathbf{Z}_2)$$

satisfies

$$\partial_q(\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_q) = \sum_{i=0}^{q} \mathbf{v}_0 \cdots \hat{\mathbf{v}}_i \cdots \mathbf{v}_q,$$

that is, $\partial_q$ sends the boundary of a q-simplex to the sum of all of its (q – 1)-dimensional faces. Proving $\partial_{q-1} \circ \partial_q = 0$ is easy in this case because each (q – 2)-dimensional face a q-simplex belongs to precisely two (q – 1)-dimensional faces of the simplex and 2 = 0 in $\mathbf{Z}_2$. The *mod 2 homology groups* of K, $H_q(K;\mathbf{Z}_2)$, are now defined to be the usual quotient group of the kernel of $\partial_q$ (called the *mod 2 q-cycles*) by the image of $\partial_{q+1}$ (called the *mod 2 q-boundaries*).

We can relate the mod 2 q-chains of K to subsets of $|K|$.

**Definition.** Let $c \in C_q(K;G)$. The *support* of c, denoted by $|c|$, is the union of all the q-simplices appearing in c with a nonzero coefficient.

Observe that if we represent q-chains c in $C_q(K;\mathbf{Z}_2)$ by the subset $|c|$ of $|K|$, which is their support, then

(1) The subset $|c + d|$ is the closure of the symmetric difference $|c| \Delta |d|$ of the subsets $|c|$ and $|d|$.
(2) The subset $|\partial_q(c)|$ is the union of all the (q – 1)-simplices that are the face of an odd number of q-simplices appearing in c.

Now, since $\mathbf{Z}_2$ is a field, the groups $C_q(K;\mathbf{Z}_2)$, $Z_q(K;\mathbf{Z}_2)$, $B_q(K;\mathbf{Z}_2)$, and $H_q(K;\mathbf{Z}_2)$ are actually vector spaces over $\mathbf{Z}_2$.

**Definition.** The *qth connectivity number of K*, $\kappa_q(K)$, is defined to be the dimension of the vector space $H_q(K;\mathbf{Z}_2)$ over $\mathbf{Z}_2$. If $\mathbf{X}$ is a polyhedron, then the *qth connectivity number of* $\mathbf{X}$, $\kappa_q(\mathbf{X})$, is defined to be the qth connectivity number of any simplicial complex K that triangulates $\mathbf{X}$.

Connectivity numbers are the mod 2 analogs of Betti numbers. They are well defined in the case of a polyhedron because the groups $H_q(\mathbf{X};\mathbf{Z}_2)$ are well defined up to isomorphism.

**7.2.6.1. Theorem.** Let K be a simplicial complex. Then

$$\chi(K) = \sum_{q=0}^{\dim K} (-1)^q \kappa_q(K).$$

**Proof.** The proof of this theorem is the same as that of Theorem 7.2.3.10. The only difference is that one uses the mod 2 groups now and works with the dimensions of these vector spaces rather than the ranks of the corresponding groups in the usual homology theory with integer coefficients. One also needs to use the fact that

$$\dim\ C_q(K;\mathbf{Z}_2) = n_q(K).$$

**7.2.6.2. Corollary.** If **X** is a polyhedron, then

$$\chi(\mathbf{X}) = \sum_{q=0}^{\dim X} (-1)^q \kappa_q(\mathbf{X}).$$

Connectivity numbers have a simple geometric interpretation in the special case of surfaces.

**7.2.6.3. Lemma.** Let **S** be a closed and compact combinatorial surface. Then

(1) $\kappa_0(\mathbf{S}) = 1$.
(2) $\kappa_2(\mathbf{S}) = 1$.
(3) $\kappa_1(\mathbf{S}) = 2 - \chi(\mathbf{S})$.

**Proof.** Exercise 7.2.6.1(b) proves (1). Next, it is easy to see that $H_2(\mathbf{S};\mathbf{Z}_2) \approx \mathbf{Z}_2$ because the sum of all the 2-simplices of **S** is a mod 2 2-cycle that generates $H_2(\mathbf{S};\mathbf{Z}_2)$. (We should point out that this fact is actually a special case of Theorem 7.5.1(1) in Section 7.5.) This proves (2). Parts (1), (2), and Corollary 7.2.6.2 now imply (3).

Lemma 7.2.6.3 shows that the first connectivity number $\kappa_1(\mathbf{S})$ (which is the only one that is interesting for surfaces) does not depend on the orientability of the surface **S**. This was certainly not the case with the first Betti number $\beta_1(\mathbf{S})$ and more evidence that mod 2 homology theory does not detect orientability properties of spaces. We shall also see this later with respect to the top dimensional homology groups of pseudomanifolds. The next theorem is the main result that we are after right now.

**7.2.6.4. Theorem.** The first connectivity number $\kappa_1(\mathbf{S})$ of a combinatorial surface **S** equals the maximum number of distinct, but not necessarily disjoint, simple closed curves in **S** along which one can cut and still have a connected set left over.

**Outline of Proof.** First of all, we need to explain the expression "simple closed curves" in our current context. A collection of subsets $\mathbf{X}_i$ of **S** will be called a collection of *simple closed curves* in **S** if each $\mathbf{X}_i$ is homeomorphic to the circle $\mathbf{S}^1$ and there

is a triangulation $(K,\varphi)$ of $\mathbf{S}$ and subcomplexes $L_i$ of K such that $\varphi(|L_i|) = \mathbf{X}_i$. We shall use this notation in the argument below.

Define a number $\delta(\mathbf{S})$ by

$$\delta(\mathbf{S}) = \max\{k \mid \mathbf{S} - (\mathbf{X}_1 \cup \mathbf{X}_2 \cup \ldots \cup \mathbf{X}_k) \text{ is connected, where the } \mathbf{X}_i \text{ are distinct simple closed curves in } \mathbf{S}\}.$$

We need to prove that $\kappa_1(\mathbf{S}) = \delta(\mathbf{S})$. It is easy to see from the normal form for a surface that was given in Chapter 6 and Lemma 7.2.6.3(3) that

$$\delta(\mathbf{S}) \geq \kappa_1(\mathbf{S}) = 2 - \chi(\mathbf{S}). \tag{7.6}$$

Conversely, let $\delta(\mathbf{S}) = k$ and assume that $k > \kappa_1(\mathbf{S})$. Let $\Sigma_i$ be the generator of $H_1(L_i;\mathbf{Z}_2) = Z_1(L_i;\mathbf{Z}_2) \subseteq Z_1(K,\mathbf{Z}_2)$, that is, $\Sigma_i$ is the sum of the 1-simplices in $L_i$. Since $\kappa_1(\mathbf{S})$ is the dimension of the vector space $H_1(K;\mathbf{Z}_2)$ and $k > \kappa_1(\mathbf{S})$, the 1-cycles $\Sigma_i$ determine a linearly dependent set of elements in $H_1(K;\mathbf{Z}_2)$. Therefore, there must be a 2-chain $c \in C_2(K;\mathbf{Z}_2)$, so that

$$\partial_2(c) = a_1\Sigma_1 + a_2\Sigma_2 + \cdots + a_k\Sigma_k,$$

where $a_i \in \{0,1\}$ and not all $a_i$ are zero. The chain c cannnot be the sum of all the 2-simplices of K, because it would be easy to check that $\partial_2(c) = 0$ in that case. Since we are assuming that $\partial_2(c) \neq 0$, at least one 2-simplex $\boldsymbol{\sigma}$ of K does not belong to c. Using this fact one shows that

$$S - (\mathbf{X}_1 \cup \mathbf{X}_2 \cup \cdots \cup \mathbf{X}_k)$$

is not connected. This contradicts our initial hypothesis and proves that $k > \kappa_1(\mathbf{S})$ is impossible, and so

$$\delta(\mathbf{S}) \leq \kappa_1(\mathbf{S}). \tag{7.7}$$

Inequalities (7.6) and (7.7) prove the theorem.

Finally, the mod 2 homology groups can be computed using "mod 2" incidence matrices.

## 7.3 Cohomology Groups

In addition to homology groups there are also cohomology groups associated to a space. These groups are a kind of dual of the homology groups. We shall run into them in Section 7.5.2 when we discuss the Poincaré duality theorem for manifolds. They provide a formal setting that makes proving facts about homology easier, even though they are closely related to the latter and add nothing new as far as the **group** structures are concerned. However, it is possible to define a natural **ring** structure for

them whose analog on the homology level is more complicated. This richer algebraic structure leads to a whole host of new topological invariants. We outline the definition of cohomology groups in the case of simplicial complexes.

Let K be a simplicial complex.

**Definition.** The *qth cochain group* of K, denoted by $C^q(K)$, is defined by

$$C^q(K) = \mathrm{Hom}\,(C_q(K), \mathbf{Z}).$$

Define the *coboundary map*

$$\delta^q : C^q(K) \to C^{q+1}(K)$$

by

$$\left(\delta^q(f)\right)(c) = f\,(\partial_q\,(c)), \quad \text{for } f \in C^q(K) \text{ and } c \in C_{q+1}(K).$$

Using the fact that $\partial_{q-1} \circ \partial_q = 0$, it is straightforward to check that $\delta^{q+1} \circ \delta^q = 0$, so that $\mathrm{im}\ \delta^{q-1} \subseteq \ker \delta^q$. The elements of $\mathrm{im}\ \delta^{q-1}$ and $\ker \delta^q$ are called *q-coboundaries* and *q-cocycles*, respectively.

**Definition.** The *qth cohomology group* of K, denoted by $H^q(K)$, is defined by

$$H^q(K) = \frac{\ker \delta^q}{\mathrm{im}\ \delta^{q-1}}.$$

**Definition.** Let K and L be simplicial complexes and $f: K \to L$ a simplicial map. Define the *induced homomorphism*

$$f^{\#q}: C^q(L) \to C^q(K)$$

by

$$\left(f^{\#q}(g)\right)(c) = g(f_{\#q}(c)), \quad \text{for } g \in C^q(L) \text{ and } c \in C_q(K).$$

It is easy to show that

$$\delta^q \circ f^{\#q} = f^{\#q+1} \circ \delta^q,$$

and so $f^{\#q}$ induces well-defined *induced homomorphisms*

$$f^{*q} : H^q(L) \to H^q(K).$$

We mentioned earlier that cohomology groups do not add anything new as groups. The following theorem is one example of this.

**7.3.1. Theorem.** Let K be a simplicial complex. If $T_q(K)$ is the torsion subgroup of $H_q(K)$ and if $r_q$ is the rank of $H_q(K)$, then

$$H^q(K) \approx T_{q-1}(K) \oplus (\text{free abelian group of rank } r_q).$$

**Proof.** See [Cair68].

Another point made earlier was that one important difference between cohomology and homology is that cohomology admits a natural product. Here is how one gets this product. First, order the vertices of the simplicial complex K and let "<" denote this ordering.

**Definition.** If $f \in C^p(K)$ and $g \in C^q(K)$, then define $f \cdot g \in C^{p+q}(K)$ by

$$(f \cdot g)([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_{p+q}]) = f([\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_p])g([\mathbf{v}_0\mathbf{v}_{p+1} \cdots \mathbf{v}_{p+q}])$$

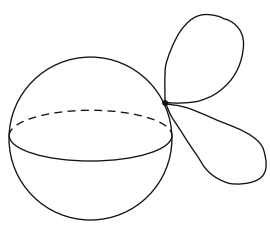
for all oriented $(p+q)$-simplices $[\mathbf{v}_0\mathbf{v}_1 \ldots \mathbf{v}_{p+q}]$ of K with $\mathbf{v}_0 < \mathbf{v}_1 < \ldots \mathbf{v}_{p+q}$. This product of cochains induces a product

$$\cup : H^p(K) \times H^q(K) \to H^{p+q}(K)$$

called the *cup product*.

Two distinct orderings of the vertices of K will induce isomorphic product structures on the cohomology groups. The cup product makes the cohomology groups into a "graded ring." As an example of how the cohomology ring gives more information, consider the space **X** in Figure 7.15 that consists of the wedge of a sphere and two circles. One can show that **X** has the same homology groups as the torus $\mathbf{S}^1 \times \mathbf{S}^1$ (Exercise 7.3.1), so that homology cannot tell those two spaces apart. On the other hand, the cohomology ring structure of **X** and the torus are different (even though both have the same cohomology **groups**). By the way, **X** does **not** have the same homotopy type as the torus.

This concludes our brief overview of cohomology groups, but we shall get more glimpses of them in the future.



$X = S^2 \vee S^1 \vee S^1$

**Figure 7.15.** A space with the same homology groups as the torus.

# 7.4 Homotopy Theory

## 7.4.1 The Fundamental Group

We motivated homology theory by saying that it was an attempt to analyze the "holes" in a space. For homology theory "holes" were treated algebraically, but we pointed out at the time that the more natural classification of holes would be via homotopy theory. We shall now take up this homotopy approach, beginning with the one-dimensional "holes" or closed paths. By studying the number of homotopy classes of closed paths we shall arrive at another important invariant associated to a topological space. However, rather than studying maps of the circle it is convenient to use maps of an interval where both end points are mapped to the same point because this simplifies a number of formulas. Such maps are clearly equivalent to maps of a circle. Throughout this section **I** will, as usual, denote the unit interval [0,1].

Let **X** be a pointed topological space with base point $\mathbf{x}_0$. We define a composition or "product" of closed paths in **X**.

**Definition.** Given maps $\alpha, \beta : (\mathbf{I}, \partial\mathbf{I}) \to (\mathbf{X}, \mathbf{x}_0)$ define a map $\alpha * \beta : (\mathbf{I}, \partial\mathbf{I}) \to (\mathbf{X}, \mathbf{x}_0)$ by

$$(\alpha * \beta)(t) = \alpha(2t) \quad , \text{if } 0 \le t \le \frac{1}{2},$$
$$= \beta(2t-1), \text{ if } \frac{1}{2} \le t \le 1.$$

See Figure 7.16. If one thinks of $\alpha$ and $\beta$ as describing paths that one walks along, then $\alpha * \beta$ corresponds to first walking along $\alpha$ at twice the original speed and then along $\beta$, also at twice the speed. The next four lemmas lead up to the main result, which is Theorem 7.4.1.5.

**7.4.1.1. Lemma.** Let $\alpha, \alpha', \beta, \beta' : (\mathbf{I}, \partial\mathbf{I}) \to (\mathbf{X}, \mathbf{x}_0)$. If $\alpha \simeq_{\partial\mathbf{I}} \alpha'$ and $\beta \simeq_{\partial\mathbf{I}} \beta'$, then $\alpha * \beta \simeq_{\partial\mathbf{I}} \alpha' * \beta'$.

**Proof.** If f(t,s) and g(t,s) are homotopies between $\alpha, \alpha'$ and $\beta, \beta'$, respectively, then the map

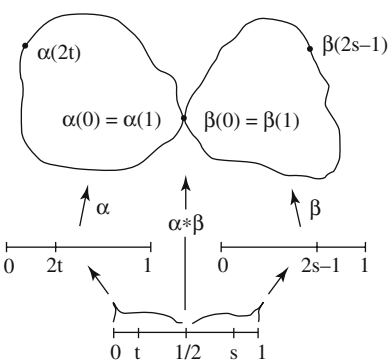
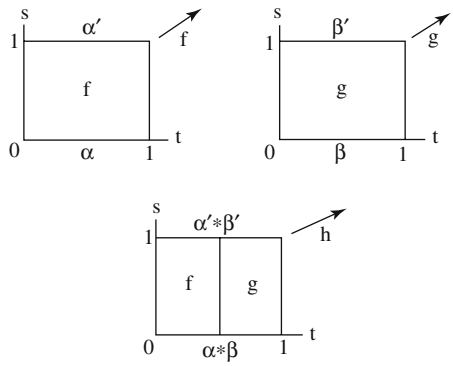



**Figure 7.16.** The composition of closed paths.

**Figure 7.17.** The homotopies in Lemma 7.4.1.2.




$$h:\ (\mathbf{I}\times\mathbf{I}, 0\times\partial\mathbf{I}\cup 1\times\partial\mathbf{I})\to(\mathbf{X},\mathbf{x}_0)$$

defined by

$$\begin{aligned} h(t,s) &= f(2t,s), && \text{if } 0\le t\le \frac{1}{2},\\ &= g(2t-1,s), && \text{if } \frac{1}{2}\le t\le 1, \end{aligned}$$

is a homotopy between $\alpha*\beta$ and $\alpha'*\beta'$. See Figure 7.17.

**7.4.1.2. Lemma.** Given maps $\alpha$, $\beta$, $\gamma:(\mathbf{I},\partial\mathbf{I})\to(\mathbf{X},\mathbf{x}_0)$, then $(\alpha*\beta)*\gamma \simeq_{\partial\mathbf{I}} \alpha*(\beta*\gamma)$.

**Proof.** It is easy to check that

$$\begin{aligned} ((\alpha*\beta)*\gamma)(t) &= \alpha(4t), && \text{for } 0\le t\le \frac{1}{4},\\ &= \beta(4t-1), && \text{for } \frac{1}{4}\le t\le \frac{1}{2},\\ &= \gamma(2t-1), && \text{for } \frac{1}{2}\le t\le 1, \end{aligned}$$

and

$$\begin{aligned} (\alpha*(\beta*\gamma))(t) &= \alpha(2t), && \text{for } 0\le t\le \frac{1}{2},\\ &= \beta(4t-2), && \text{for } \frac{1}{2}\le t\le \frac{3}{4},\\ &= \gamma(4t-3), && \text{for } \frac{3}{4}\le t\le 1. \end{aligned}$$

The map h defined by

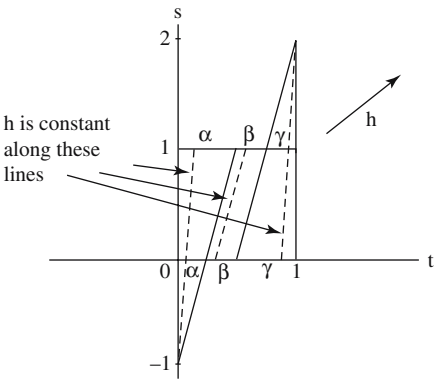



**Figure 7.18.** Associativity of closed path composition.

$$\begin{aligned} h(t,s) &= \alpha\left(\frac{4t}{s+1}\right), && \text{for } 0 \le t \le \frac{s+1}{4}, \\ &= \beta(4t-s-1), && \text{for } \frac{s+1}{4} \le t \le \frac{s+2}{4}, \\ &= \gamma\left(\frac{4t-s-2}{2-s}\right), && \text{for } \frac{s+2}{4} \le t \le 1, \end{aligned}$$

is a homotopy between $(\alpha * \beta) * \gamma$ and $\alpha * (\beta * \gamma)$. See Figure 7.18.

Define the constant path $c : \mathbf{I} \to \mathbf{X}$ by $c(t) = \mathbf{x}_0$.

**7.4.1.3. Lemma.** For any $\alpha : (\mathbf{I}, \partial\mathbf{I}) \to (\mathbf{X}, \mathbf{x}_0)$, $\alpha * c \simeq_{\partial\mathbf{I}} \alpha \simeq_{\partial\mathbf{I}} c * \alpha$.

**Proof.** Define a map h by

$$\begin{aligned} h(t,s) &= \alpha\left(\frac{2t}{s+1}\right), && \text{for } 0 \le t \le \frac{s+1}{2}, \\ &= \mathbf{x}_0, && \text{for } \frac{s+1}{2} \le t \le 1. \end{aligned}$$

Then h is a homotopy between $\alpha * c$ and $\alpha$. See Figure 7.19. A similar homotopy can be defined between $c * \alpha$ and $\alpha$, proving the lemma.

**7.4.1.4. Lemma.** Given $\alpha$, define $\beta$ by $\beta(t) = \alpha(1 - t)$. Then $\alpha * \beta \simeq_{\partial\mathbf{I}} c \simeq_{\partial\mathbf{I}} \beta * \alpha$.

**Proof.** First observe that

$$\begin{aligned} (\alpha * \beta)(t) &= \alpha(2t), && \text{for } 0 \le t \le \frac{1}{2}, \\ &= \alpha(2-2t), && \text{for } \frac{1}{2} \le t \le 1. \end{aligned}$$

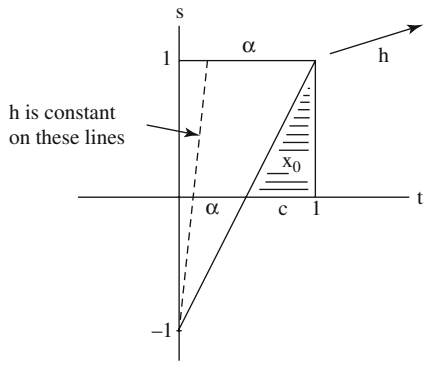



**Figure 7.19.** The constant path acts as an identity.

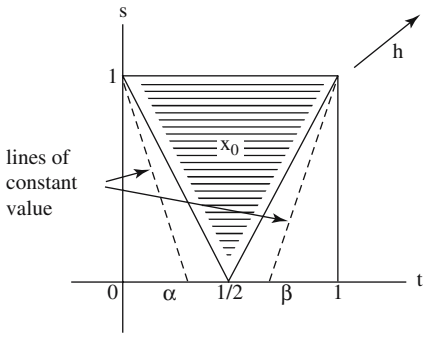



**Figure 7.20.** Reverse traversal gives the inverse.

It is easy to check that the map h defined by

$$\begin{aligned} h(t,s) &= \alpha(2ts), && \text{for } 0 \le t \le \frac{1}{2}, \\ &= \alpha(2s - 2ts), && \text{for } \frac{1}{2} \le t \le 1, \end{aligned}$$

is a homotopy between $\alpha * \beta$ and the constant path c. See Figure 7.20. A similar map shows that $\beta * \alpha$ is homotopic to c.

Let $\pi_1(\mathbf{X},\mathbf{x}_0)$ denote the set of equivalence classes of maps $\alpha:(\mathbf{I},\partial\mathbf{I}) \to (\mathbf{X},\mathbf{x}_0)$ with respect to the equivalence relation $\simeq_{\partial\mathbf{I}}$. More precisely,

$$\pi_1(\mathbf{X},\mathbf{x}_0) = [(\mathbf{I},\partial\mathbf{I}),(\mathbf{X},\mathbf{x}_0)].$$

Define a product $*$ on $\pi_1(\mathbf{X},\mathbf{x}_0)$ as follows: If $[\alpha]$, $[\beta] \in \pi_1(\mathbf{X},\mathbf{x}_0)$, then $[\alpha]*[\beta] = [\alpha*\beta]$.

**7.4.1.5. Theorem.** The operation $*$ on $\pi_1(\mathbf{X},\mathbf{x}_0)$ is well defined and makes $\pi_1(\mathbf{X},\mathbf{x}_0)$ into a group.

**Proof.** The fact that $*$ is well defined follows from Lemma 7.4.1.1. Lemma 7.4.1.2 shows that $*$ is associative. Lemma 7.4.1.3 shows that if c(t) is the constant path, then

[c] is the identity for $*$. Lemma 7.4.1.4 shows that each element of $\pi_1(\mathbf{X},\mathbf{x}_0)$ has an inverse with respect to $*$.

**Definition.** The group $\pi_1(\mathbf{X},\mathbf{x}_0)$ is called the *fundamental group* or *first homotopy group* of the pointed space $(\mathbf{X},\mathbf{x}_0)$. The point $\mathbf{x}_0$ is called the *base point* for this fundamental group.

So now we have another group associated to a space, but what information does it give us? Before we compute the group for some spaces we look at some general properties of it. First of all, the isomorphism type of the fundamental group is independent of the base point if the space is path-connected.

**7.4.1.6. Theorem.** If $\mathbf{X}$ is path-connected, then $\pi_1(\mathbf{X},\mathbf{x}_0)$ is isomorphic to $\pi_1(\mathbf{X},\mathbf{x}_1)$ for all $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbf{X}$.

**Proof.** Let $\gamma: \mathbf{I} \to \mathbf{X}$ be a path from $\mathbf{x}_1$ to $\mathbf{x}_0$. Define a map

$$T: \ \pi_1(\mathbf{X},\mathbf{x}_0) \to \pi_1(\mathbf{X},\mathbf{x}_1)$$

by

$$T([\alpha]) = [\alpha_\gamma],$$

where $\alpha_\gamma : (\mathbf{I},\partial\mathbf{I}) \to \mathbf{X}$ is defined by

$$\begin{aligned} \alpha_\gamma(t) &= \gamma(3t), && \text{for } t \in \left[0,\frac{1}{3}\right], \\ &= \alpha(3t-1), && \text{for } t \in \left[\frac{1}{3},\frac{2}{3}\right], \\ &= \gamma(3-3t), && \text{for } t \in \left[\frac{2}{3},1\right]. \end{aligned}$$

The map $\alpha_\gamma(t)$ is the path that walks along $\gamma(t)$, then $\alpha(t)$, and then backtracks along $\gamma(t)$. See Figure 7.21. It is easy to check that T is a well-defined isomorphism (Exercise 7.4.1.1).

**Note.** Because of Theorem 7.4.1.6, the base point is often omitted for path-connected spaces $\mathbf{X}$ and $\pi_1(\mathbf{X})$ is used to denote $\pi_1(\mathbf{X},\mathbf{x}_0)$ for some $\mathbf{x}_0 \in \mathbf{X}$.

**7.4.1.7. Theorem.** A contractible space has a trivial fundamental group.

**Proof.** The homotopy that shows the space is contractible to a point easily provides a homotopy between every closed path in the space with the constant path (Exercise 7.4.1.2).

**7.4.1.8. Corollary.** A point, $\mathbf{R}^n$, and $\mathbf{D}^n$ all have a trivial fundamental group.

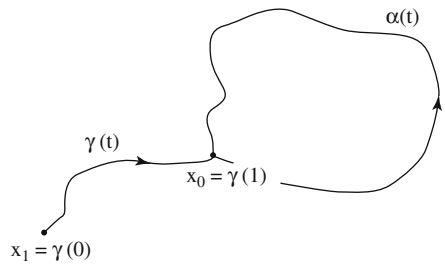



**Figure 7.21.** Proving that the fundamental group is independent of the base point.

**7.4.1.9. Theorem.** The fundamental group of $\mathbf{S}^2$ is trivial.

**Proof.** Here is a sketch of the proof. Let $\alpha:(\mathbf{I},\partial\mathbf{I}) \to (\mathbf{S}^2,\mathbf{e}_1)$. The Simplicial Approximation Theorem implies that $\alpha$ is homotopic (relative to $\partial\mathbf{I}$) to a map that misses a point $\mathbf{p} \neq \mathbf{e}_1$ in $\mathbf{S}^2$. Since $\mathbf{S}^2 - \mathbf{p}$ can be contracted to $\mathbf{e}_1$ this proves that a is homotopic to the constant map and the theorem is proved.

Just in case the reader is beginning to think that the fundamental group is always trivial, we give some examples of simple spaces for which the group is nontrivial.

**7.4.1.10. Theorem.**

(1) $\pi_1(\mathbf{S}^1) \approx \mathbf{Z}$.
(2) $\pi_1(\mathbf{P}^2) \approx \mathbf{Z}_2$.
(3) The fundamental group of a wedge of two circles (figure eight) is a free group on two generators.

**Proof.** See [Mass67] or [Cair68]. One way to prove part (1) is to show that the isomorphism is defined by the degree of the map as sketched in Section 5.7 and defined rigorously in Section 7.5.1. Later in Corollary 7.4.2.23 we shall see alternate proofs of (1) and (2).

Next, we look at how the fundamental group behaves with respect to continuous maps. Let $(\mathbf{X},\mathbf{x}_0)$ and $(\mathbf{Y},\mathbf{y}_0)$ be pointed spaces. Let $f:(\mathbf{X},\mathbf{x}_0) \to (\mathbf{Y},\mathbf{y}_0)$ be a continuous map. Define

$$f_*: \pi_1(\mathbf{X},\mathbf{x}_0) \to \pi_1(\mathbf{Y},\mathbf{y}_0)$$

by

$$f_*([\alpha]) = [f\circ\alpha].$$

**7.4.1.11. Lemma.** The map $f_*$ is a well-defined homomorphism of groups.

**Proof.** This is Exercise 7.4.1.3.

**Definition.** The homomorphism $f_*$ is called the homomorphism *induced* by the continuous map f.

**7.4.1.12. Theorem.**

(1) If $f, g:(\mathbf{X},\mathbf{x}_0) \to (\mathbf{Y},\mathbf{y}_0)$ are homotopic continuous maps, then

$$f_* = g_* : \pi_1(\mathbf{X},\mathbf{x}_0) \to \pi_1(\mathbf{Y},\mathbf{y}_0).$$

(2) If $f:(\mathbf{X},\mathbf{x}_0) \to (\mathbf{Y},\mathbf{y}_0)$ and $g:(\mathbf{Y},\mathbf{y}_0) \to (\mathbf{Z},\mathbf{z}_0)$ are continuous maps, then

$$g_* \circ f_* = (g \circ f)_* : \pi_1(\mathbf{X},\mathbf{x}_0) \to \pi_1(\mathbf{Z},\mathbf{z}_0).$$

**Proof.** This is Exercise 7.4.1.4.

We can now prove the homotopy invariance of the fundamental group.

**7.4.1.13. Theorem.** Homotopy equivalent spaces have isomorphic fundamental groups.

**Proof.** This is an easy consequence of Theorem 7.4.1.12.

Note that Theorem 7.4.1.7 is actually an easy consequence of Theorem 7.4.1.13 since a contractible space has the same homotopy type as a point.

There is a nice relationship between the fundamental group of two spaces and that of their product.

**7.4.1.14. Theorem.** Let $(\mathbf{X},\mathbf{x}_0)$ and $(\mathbf{Y},\mathbf{y}_0)$ be pointed spaces and let

$$p: \mathbf{X} \times \mathbf{Y} \to \mathbf{X} \quad \text{and} \quad q: \mathbf{X} \times \mathbf{Y} \to \mathbf{Y}$$

be the natural projections defined by $p(\mathbf{x},\mathbf{y}) = \mathbf{x}$ and $q(\mathbf{x},\mathbf{y}) = \mathbf{y}$. Then the map

$$\sigma: \pi_1(\mathbf{X} \times \mathbf{Y}, \mathbf{x}_0 \times \mathbf{y}_0) \to \pi_1(\mathbf{X},\mathbf{x}_0) \times \pi_1(\mathbf{Y},\mathbf{y}_0)$$

defined by

$$\sigma([\alpha]) = ([p \circ \alpha],[q \circ \alpha]),$$

is an isomorphism.

**Proof.** See [Mass67].

Theorem 7.4.1.14 enables us to compute many more fundamental groups.

**7.4.1.15. Corollary.** $\pi_1(\mathbf{S}^1 \times \mathbf{S}^1) \approx \mathbf{Z} \oplus \mathbf{Z}$.

**Proof.** The corollary follows from Theorems 7.4.1.10(1) and 7.4.1.14. See also Corollary 7.4.2.23(3).

Theorem 7.4.1.10(3) showed that the fundamental group of a space is not necessarily abelian. Is there any connection between it and the first homology group? After all, in both cases we are dealing with one-dimensional "holes." To answer that question we describe a natural map from one to the other.

**Note.** To simplify the discussion below we are pretending (as we earlier said we would) that polyhedra have well-defined homology groups.

Let $\mathbf{X}$ be a connected polyhedron and let $\mathbf{x}_0 \in \mathbf{X}$. Define

$$\mu : \pi_1(\mathbf{X},\mathbf{x}_0) \to H_1(\mathbf{X})$$

as follows: Let $[\alpha] \in \pi_1(\mathbf{X},\mathbf{x}_0)$, where $\alpha : (\mathbf{I},\partial\mathbf{I}) \to (\mathbf{X},\mathbf{x}_0)$. If $\varphi : \mathbf{I} \to \mathbf{S}^1$ is the map

$$\varphi(t) = (\cos 2\pi t, \sin 2\pi t),$$

then $\alpha$ induces a unique map $\beta : \mathbf{S}^1 \to \mathbf{X}$ with the property that $\alpha(t) = \beta(\varphi(t))$. ($\beta$ is the unique map that makes the diagram

$$\begin{tikzcd} S^1 \arrow[r, "\beta"] & X \\ (I,\partial I) \arrow[u, "\varphi"] \arrow[ur, "\alpha"'] & \end{tikzcd}$$

commutative.) Now $\beta$ induces a map $\beta_*$ on homology groups. If $\iota$ is a fixed ("standard") generator of $H_1(\mathbf{S}^1)$, then

$$\mu([\alpha]) = \beta_*(\iota) \in H_1(\mathbf{X}).$$

**7.4.1.16. Theorem.**

(1) The map $\mu$ defines a homomorphism of groups called the *Hurewicz homomorphism*.
(2) The map $\mu$ sends $\pi_1(\mathbf{X},\mathbf{x}_0)$ onto $H_1(\mathbf{X})$.
(3) The kernel of $\mu$ is the commutator subgroup of $\pi_1(\mathbf{X},\mathbf{x}_0)$.

**Proof.** See [Cair68].

It follows from Theorem 7.4.1.16 that $H_1(\mathbf{X})$ is the abelianization of $\pi_1(\mathbf{X},\mathbf{x}_0)$. The advantage of the fundamental group of a space is that it gives somewhat more information about the space than the first homology group. The disadvantage is that it is more complicated to compute.

**Definition.** A path-connected space that has a trivial fundamental group is called *simply connected*.

Analyzing spaces and maps gets much more complicated if the spaces involved are not simply connected. The fundamental group has a subtle but significant influence on the topology of a space and its name is quite appropriate. It is probably the single most important group from the point of view of algebraic topology. A great many theorems have as part of their hypotheses the assumption that a space is simply connected. See for example Theorems 7.4.3.7 and results in Section 8.7, 8.9, and 8.10.

A nice way to summarize some aspects of the fundamental group and its relationship to the first homology group is as follows: If

$$f: (\mathbf{X}, \mathbf{x}_0) \to (\mathbf{Y}, \mathbf{y}_0)$$

is a continuous map, then (again ignoring the current nonuniqueness of homology groups) there is a commutative diagram

$$\begin{array}{ccc} \pi_1(\mathbf{X}, \mathbf{x}_0) & \xrightarrow[f_*]{} & \pi_1(\mathbf{Y}, \mathbf{y}_0) \\ \mu\downarrow & & \downarrow\mu \\ H_1(\mathbf{X}) & \xrightarrow[f_*]{} & H_1(\mathbf{Y}) \end{array},$$

where $\mu$ is the Hurewicz homomorphism.

We end this section with an application of the fundamental group. The group plays a central role in the study of knots. Some references for knot theory are [CroF65], [Livi93], [Rolf76], and [Mass67].

**Definition.** A subspace $\mathbf{K}$ of $\mathbf{R}^3$ is called a *knot* if $\mathbf{K}$ is homeomorphic to $\mathbf{S}^1$. The space $\mathbf{R}^3 - \mathbf{K}$ is called the *complement* of the knot $\mathbf{K}$. Two knots $\mathbf{K}_1$ and $\mathbf{K}_2$ are said to be *equivalent* if there is a homeomorphism $h: \mathbf{R}^3 \to \mathbf{R}^3$, so that $h(\mathbf{K}_1) = \mathbf{K}_2$.The equivalence class of a knot is called its *knot type*. A knot is *trivial* if it is equivalent to the standard $\mathbf{S}^1$ in $\mathbf{R}^3$. A knot is called a *polygonal knot* if it is the union of a finite number of (linear) segments, that is, it is a polygonal curve. A knot is said to be *tame* if it is equivalent to a polygonal knot.

We are sticking to the traditional theory here, because the definition of a knot is sometimes generalized to include imbeddings of n-spheres, $n \geq 1$, in a space. We also need to point out that there are other variations of the definition of a knot in the literature. Sometimes knots are defined to be **maps**, that is, imbeddings $k: \mathbf{S}^1 \to \mathbf{R}^3$, rather than subsets. In that case, the equivalence of knots is defined in terms of isotopies. (Two imbeddings $h_0$ and $h_1$ are said to be *isotopic* if there exists a one-parameter family of imbeddings $h_t$, or *isotopy*, between them.) Fortunately, there is not much difference between the theories. For example, if we stick to orientation-preserving homeomorphisms, then two knots are equivalent using our definition if and only if they are isotopic. (We shall define what it means for a homeomorphism between oriented manifolds to be orientation preserving in Section 7.5.1. A homeomorphism $h: \mathbf{R}^3 - \mathbf{K} \to \mathbf{R}^3 - \mathbf{K}$ is said to be *orientation preserving* if its extension to $\mathbf{S}^3 \to \mathbf{S}^3$ is.)

Sometimes it is convenient to consider knots in $\mathbf{S}^3$ rather than $\mathbf{R}^3$ because $\mathbf{S}^3$ is a compact space, but there is again no real difference in the theory since, using the stereographic projections, $\mathbf{S}^3$ can be thought of as just $\mathbf{R}^3$ with one point added. Note also that, since all knot, are homeomorphic to $\mathbf{S}^1$, classifying them is not a question of determining if they themselves are homeomorphic because they are. What makes knots different is their imbedding in $\mathbf{R}^3$. Every knot in the plane is necessarily trivial by the Schoenflies theorem.

In order not to have to deal with wild imbeddings, one also usually assumes that knots are polygonal.

**Definition.** Let $\mathbf{K}$ be a knot. The fundamental group $\pi_1(\mathbf{R}^3 - \mathbf{K})$ is called the *group of the knot* $\mathbf{K}$. (The base point of the fundamental group was omitted because we are only interested in the group up to isomorphism.)

The group of a knot plays a large role in the study of knots but does not determine the knot completely because there exist inequivalent knots that have the same knot group, such as for example, the *square knot* and the *granny knot* shown in Figure 7.22. Certainly, equivalent knots have isomorphic knot groups because their complements are homeomorphic. The knot group is only one of many interesting invariants associated to a knot.

Before we list a few important known facts about the classification of knots, we define a well-known infinite family of knots that serve as useful examples.

**Definition.** A *torus knot of type* $(p,q)$, where p and q are relatively prime, is a knot that can be imbedded in a torus and has the property that it cuts a meridian circle of the torus in p points and a circle of latitude in q points. In cylindrical coordinates, a specific instance of such a knot is the curve

$$r = 2 + \cos(q\theta/p)$$
$$z = \sin(q\theta/p)$$

that lies in the torus in $\mathbf{R}^3$ (the circle in the x-z plane with center (2,0,0) and radius 1 rotated about the z-axis) defined by the equation

$$(r-2)^2 + z^2 = 1.$$

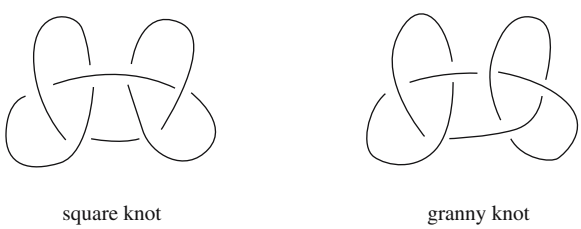



**Figure 7.22.** Two inequivalent knots with isomorphic knot groups.

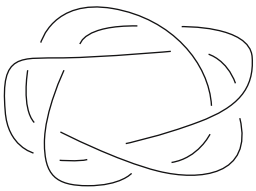


**Figure 7.23.** A torus knot of type (3,5).

Figure 7.23 shows an example of a torus knot.

**7.4.1.17. Theorem.**

(1) A tame knot is trivial if and only if the group of the knot is infinite cyclic (isomorphic to $\mathbf{Z}$). There is an algorithm that determines whether or not a knot is trivial.
(2) Two tame knots have homotopy equivalent complements if and only if their knot groups are isomorphic. (Conjecture: If two tame knots have homeomorphic complements, then they have the same knot type.)
(3) There exist infinitely many knot types. For example, the torus knots of type (p,q) are all inequivalent.
(4) The abelianization of every knot group is infinite cyclic.
(5) If $\mathbf{K}$ is a tame knot, then $\pi_i(\mathbf{R}^3 - \mathbf{K}) = 0$ for $i > 1$.

**Proof.** The proofs of most of these facts are much too complicated to give here. See the references for knot theory listed earlier.

## 7.4.2 Covering Spaces

The topic of this section is intimately connected with the fundamental group but also has important applications in other areas such as complex analysis and Riemann surfaces. Section 8.10 in the next chapter will continue the discussion and discuss the related topic of vector bundles.

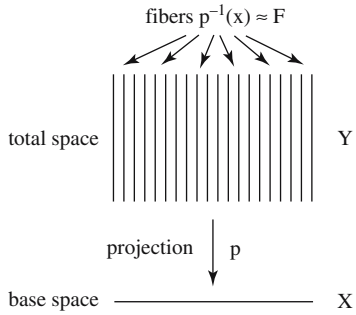
We begin with some basic terminology and motivational remarks. See Figure 7.24.

**Definition.** A *bundle over a space* $\boldsymbol{X}$ is a pair $(\mathbf{Y},p)$, where $\mathbf{Y}$ is a topological space and $p:\mathbf{Y} \to \mathbf{X}$ is a continuous surjective map. One calls $\mathbf{Y}$ the *total space*, p the *projection*, and $\mathbf{X}$ the *base space* of the bundle. The inverse images $p^{-1}(\mathbf{x}) \subseteq \mathbf{Y}$ for $\mathbf{x} \in \mathbf{X}$, are called the *fibers* of the bundle.

In our current context we should think of the total space of a bundle as consisting of a union of fibers that are glued together appropriately. Of course, the general case of an arbitrary surjective map p does not lead to anything interesting. The interesting case is where all the fibers are homeomorphic to a fixed space $\mathbf{F}$. The obvious example of that is the product of the base space and $\mathbf{F}$.

**Definition.** A bundle over $\mathbf{X}$ of the form $(\mathbf{X} \times \mathbf{F},p)$, where p is the projection onto the first factor defined by $p(\mathbf{x},\mathbf{f}) = \mathbf{x}$, is called the *product bundle with fiber F*.

**Figure 7.24.** Basic bundle terminology.




Next, we define a notion of equivalence of bundles over a space. We begin by defining general bundle maps. They should preserve the fibers (map fibers to fibers) since that is the only structure present.

**Definition.** A *bundle map* from a bundle $(\mathbf{Y}_1,p_1)$ over a space $\mathbf{X}$ to a bundle $(\mathbf{Y}_2,p_2)$ over $\mathbf{X}$ is a map

$$f:\ \mathbf{Y}_1 \to \mathbf{Y}_2$$

with the property that

$$\begin{tikzcd} \mathbf{Y}_1 \arrow[rr, "f"] \arrow[dr, "p_1"'] & & \mathbf{Y}_2 \arrow[dl, "p_2"] \\ & \mathbf{X} & \end{tikzcd}$$

is a commutative diagram ($p_1 = p_2 \circ f$). The bundle map f is called a *bundle isomorphism* and we say that the bundles $(\mathbf{Y}_1,p_1)$ and $(\mathbf{Y}_2,p_2)$ are *isomorphic* if f is a homeomorphism. If $(\mathbf{Y},p) = (\mathbf{Y}_1,p_1) = (\mathbf{Y}_2,p_2)$, then a bundle isomorphism is called a *bundle automorphism* of $(\mathbf{Y},p)$.

**Definition.** A bundle $(\mathbf{Y},p)$ over a space $\mathbf{X}$ that is isomorphic to a product bundle is called a *trivial bundle*. The bundle is called a *locally trivial bundle* if for every $\mathbf{x} \in \mathbf{X}$ there is an open neighborhood $\mathbf{U}$ of $\mathbf{x}$ in $\mathbf{X}$ such that $(p^{-1}(\mathbf{U}),p\,|\,p^{-1}(\mathbf{U}))$ is isomorphic to a trivial bundle over $\mathbf{U}$.

If all locally trivial bundles were trivial bundles, there would be no point in introducing the concept of bundle. The next example describes a very simple nontrivial bundle.

**7.4.2.1. Example.** If we consider $\mathbf{P}^n$ as the quotient space of $\mathbf{S}^n$ where antipodal points are identified and let $p:\mathbf{S}^n \to \mathbf{P}^n$ be the quotient map, then one can show that $(\mathbf{S}^n,p)$ is a locally trivial bundle over $\mathbf{P}^n$ (Exercise 7.4.2.1). Every fiber is the discrete space consisting of two points. Clearly, $(\mathbf{S}^n,p)$ is not a trivial bundle because $\mathbf{S}^n$ is connected and the trivial bundle with fibers consisting of two points would not be.

It is the bundles with discrete fibers that interest us in this section. In Section 8.10 we shall look at bundles whose fibers are vector spaces.

**Definition.** A *covering space* for a space $\mathbf{X}$ is a locally trivial bundle with base space $\mathbf{X}$ with the property that every fiber is a discrete space. The covering is called an *n-fold covering* if every fiber consists of n points. The bundle automorphisms of a covering space are called *covering transformations*.

Example 7.4.2.1 already described a 2-fold covering space. Here are some more examples.

**7.4.2.2. Example.** The map

$$p: \mathbf{R} \to \mathbf{S}^1$$

$$p(t) = (\cos t, \sin t)$$

defines a covering space $(\mathbf{R}, p)$ of $\mathbf{S}^1$ whose fibers

$$p^{-1}(t) = \{t + 2\pi n \mid n \in \mathbf{Z}\}$$

are a countable discrete set of points.

**7.4.2.3. Example.** Consider the circle $\mathbf{S}^1$ as a subset of the complex plane $\mathbf{C}$. The map

$$p: \mathbf{S}^1 \to \mathbf{S}^1$$

$$p(\mathbf{z}) = \mathbf{z}^n$$

defines a bundle $(\mathbf{S}^1, p)$ over $\mathbf{S}^1$ that is an n-fold covering space for $\mathbf{S}^1$.

**7.4.2.4. Example.** The map

$$p: \mathbf{R}^2 \to \mathbf{S}^1 \times \mathbf{S}^1 \subset \mathbf{R}^2 \times \mathbf{R}^2$$

$$p(s,t) = ((\cos t, \sin t), (\cos t, \sin t)).$$

defines a is a covering space $(\mathbf{R}^2, p)$ of the torus $\mathbf{S}^1 \times \mathbf{S}^1$.

That all the total spaces in our examples were manifolds should not be surprising.

**7.4.2.5. Theorem.** The total space of a covering space of a topological manifold is a topological manifold.

**Proof.** This is obvious from the local triviality property of the bundle.

When one works with covering spaces or bundles in general, some of the most important tools are map-lifting tools.

**Definition.** Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$ and let $\gamma:[a,b] \to \mathbf{X}$ be a continuous curve. A map $\tilde{\gamma}:[a,b] \to \mathbf{Y}$ is called a *lifting* of the curve $\gamma$ starting at $\tilde{\gamma}(a)$ if we have a commutative diagram

$$\begin{tikzcd} & \mathbf{Y} \arrow[d, "p"] \\ {[a,b]} \arrow[ur, "\tilde{\gamma}"] \arrow[r, "\gamma"] & \mathbf{X}, \end{tikzcd}$$

that is, $p \circ \tilde{\gamma} = \gamma$. More generally, given a map $f:\mathbf{Z} \to \mathbf{X}$, any map $\tilde{f}:\mathbf{Z} \to \mathbf{Y}$ is called a *lifting* of f if we have a commutative diagram

$$\begin{tikzcd} & \mathbf{Y} \arrow[d, "p"] \\ \mathbf{Z} \arrow[ur, "\tilde{f}"] \arrow[r, "f"] & \mathbf{X}, \end{tikzcd}$$

that is, $p \circ \tilde{f} = f$.

**7.4.2.6. Theorem.** (The Path-Lifting Theorem) Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$. Let $\mathbf{x}_0 \in \mathbf{X}$ and $\mathbf{y}_0 \in p^{-1}(\mathbf{x}_0)$. Then every continuous curve $\gamma:[0,1] \to \mathbf{X}$ lifts to a **unique** continuous curve $\tilde{\gamma}:[0,1] \to \mathbf{Y}$ that starts at $\mathbf{y}_0$.

**Proof.** We sketch a proof of this theorem. It will give the reader a good idea of the kind of arguments one uses with covering spaces. Figure 7.25 shows what is involved.

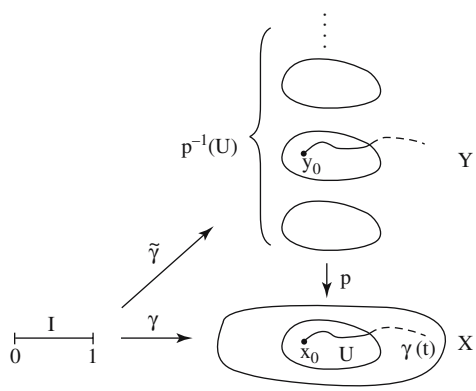



**Figure 7.25.** Lifting paths.

The easy case is where one can find an open neighborhood $\mathbf{U}$ of $\mathbf{x}_0$ over which the covering space is trivial and that contains the curve $\gamma(t)$. The set $p^{-1}(\mathbf{U})$ will consist of disjoint open sets in $\mathbf{Y}$ that are homeomorphic copies of $\mathbf{U}$. Let $\mathbf{V}$ be the one that contains $\mathbf{y}_0$ and let $p_{\mathbf{V}} = p\,|\,\mathbf{V}$. Then $\tilde{\gamma} = p_{\mathbf{V}}^{-1}\circ\gamma$ is the unique curve we seek. For the general case, we separate the proof into two parts.

We prove uniqueness first. Let $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ be two liftings of $\gamma$ that start at $\mathbf{y}_0$. Consider the sets

$$\mathbf{A} = \{t \in [0,1] \mid \tilde{\gamma}_1(t) = \tilde{\gamma}_2(t)\} \quad \text{and} \quad \mathbf{B} = \{t \in [0,1] \mid \tilde{\gamma}_1(t) \neq \tilde{\gamma}_2(t)\}.$$

These are obviously disjoint sets whose union is $[0,1]$. Using continuity, it is easy to show that both of these sets are open in $[0,1]$. Since $0 \in \mathbf{A}$, $\mathbf{A}$ is nonempty. But $[0,1]$ is connected and so $\mathbf{B}$ must be the empty set and we have proved uniqueness.

To prove the existence of a $\tilde{\gamma}$, consider the set

$$C = \{t \in [0,1] \mid \text{there is a lifting of } \gamma \text{ over} [0,t]\}.$$

Figure 7.26 should help the reader follow the rest of the argument. Since the covering space is trivial over an open neighborhood of $\mathbf{x}_0$ and we know how to lift paths over such a neighborhoods, the set $\mathbf{C}$ will contain a small neighborhood of 0 and hence, if c is the supremum of $\mathbf{C}$, then $0 < c \leq 1$. If $c = 1$, we are done. Assume that $c < 1$. Choose a neighborhood $\mathbf{U}$ of $\gamma(c)$ over which the covering space is trivial. Choose $\varepsilon > 0$, such that $[c - 2\varepsilon, c + 2\varepsilon] \subset [0,1]$ and $\gamma([c - 2\varepsilon, c + 2\varepsilon]) \subset \mathbf{U}$. By the definition of c, there is a lifting

$$\tilde{\gamma}: [0, c - \varepsilon] \to \mathbf{Y}$$

of

$$\gamma: [0, c - \varepsilon] \to \mathbf{X}.$$

Because the curve $\gamma\,|\,[c - 2\varepsilon, c + 2\varepsilon]$ lies in $\mathbf{U}$ it can be lifted to $\mathbf{Y}$, that is, the lifting $\tilde{\gamma}$ can be extended to a lifting of $\gamma\,|\,[0, c + 2\varepsilon]$. This contradicts the fact that c was the supremum of the set $\mathbf{C}$ and so $c < 1$ is impossible.

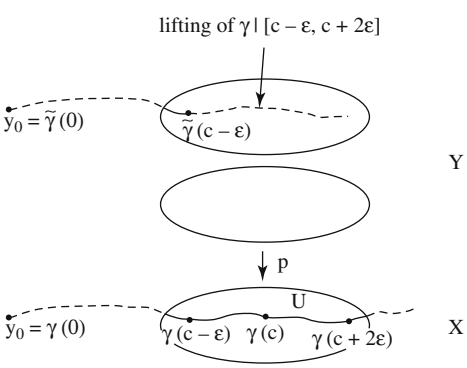
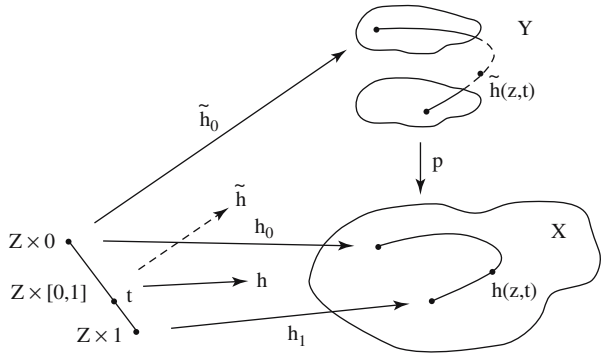



**Figure 7.26.** Proving the existence of path liftings.

**Figure 7.27.** Lifting a homotopy.




The importance of Theorem 7.4.2.6 is not only that every path in the base space lifts to path in the total space but that the lift is essentially unique, meaning that if two lifted paths agree at one point, then they agree everywhere. The unique lifting property generalizes to arbitrary connected spaces not just the interval [0,1]. Another important lifting theorem is the following:

**7.4.2.7. Theorem.** (The Homotopy Lifting Theorem) Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$. Let $h:\mathbf{Z} \times [0,1] \to \mathbf{X}$ be a continuous map. Define $h_t:\mathbf{Z} \to \mathbf{X}$ by $h_t(y) = h(y,t)$. If $\tilde{h}_0$ is a lifting of $h_0$, then h lifts to a unique continuous map $\tilde{h}:\mathbf{Z} \times [0,1] \to \mathbf{Y}$ so that $\tilde{h}(y,0) = \tilde{h}_0(y)$.

**Proof.** See [Jäni84]. Figure 7.27 tries to indicate the relationship between the various maps.

**7.4.2.8. Corollary.** (The Monodromy Lemma) Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$. Let $\gamma_0, \gamma_1:[0,1] \to \mathbf{X}$ be two continuous curves that start at the same point $\mathbf{x}_0$ and end at the same point $\mathbf{x}_1$, that is, $\mathbf{x}_0 = \gamma_0(0) = \gamma_1(0)$ and $\mathbf{x}_1 = \gamma_0(1) = \gamma_1(1)$. Assume that $\gamma_0$ and $\gamma_1$ are homotopic by a homotopy h that fixes the endpoints, that is, $h(t,0) = \mathbf{x}_0$ and $h(t,1) = \mathbf{x}_1$, for all $t \in [0,1]$. If, $\tilde{\gamma}_0, \tilde{\gamma}_1:[0,1] \to \mathbf{Y}$ are liftings of $\gamma_0$ and $\gamma_1$, respectively, that start at the same point in $\mathbf{Y}$, then $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ will end at the same point, that is, $\tilde{\gamma}_0(1) = \tilde{\gamma}_1(1)$.

**Proof.** This is an easy consequence of Theorem 7.4.2.7.

Corollary 7.4.2.8 is an important uniqueness type theorem. It says that if one lifts two homotopic paths that start and end at the same point, then the lifted paths will also end at the **same** point if they start at the same point.

The next two results describe some relationships between the fundamental groups of the total and base space of a covering space.

**7.4.2.9. Theorem.** Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$. Let $\mathbf{x}_0 \in \mathbf{X}$ and $\mathbf{y}_0 \in p^{-1}(\mathbf{x}_0)$. Then the induced homomorphism

$$p_*: \pi_1(\mathbf{Y},\mathbf{y}_0) \to \pi_1(\mathbf{X},\mathbf{x}_0)$$

is one-to-one.

**Proof.** See [Mass67] or [Jäni84]. Basically, if an element [f] maps to 0, then the map $p \circ f$ is homotopic to a constant in **X** and this homotopy lifts to a homotopy between f and the constant map in **Y**.

A natural question is if $\mathbf{y}_1 \in p^{-1}(\mathbf{x}_0)$, then what is the relation between the subgroups $p_*(\pi_1(\mathbf{Y},\mathbf{y}_1))$ and $p_*(\pi_1(\mathbf{Y},\mathbf{y}_0))$ in $\pi_1(\mathbf{X},\mathbf{x}_0)$? There is an easy answer.

**7.4.2.10. Theorem.** Let (**Y**,p) be a covering space for a space **X** and let $\mathbf{x}_0 \in \mathbf{X}$. If **Y** is connected, then the subgroups $p_*(\pi_1(\mathbf{Y},\mathbf{y}_0))$ in $\pi_1(\mathbf{X},\mathbf{x}_0)$ as $\mathbf{y}_0$ ranges over the points in $p^{-1}(\mathbf{x}_0)$ generate a conjugacy class of subgroups in $\pi_1(\mathbf{X},\mathbf{x}_0)$.

**Proof.** See [Mass67]. The result follows easily from the following observations. Let $\mathbf{y}_0, \mathbf{y}_1 \in p^{-1}(\mathbf{x}_0)$. Let $\tilde{\alpha} : [0,1] \to \mathbf{Y}$ be a curve with $\tilde{\alpha}(0) = \mathbf{y}_0$ and $\tilde{\alpha}(1) = \mathbf{y}_1$. The curve $\alpha = p \circ \tilde{\alpha} : [0,1] \to \mathbf{X}$ is a loop at $\mathbf{x}_0$. If $[\tilde{\gamma}] \in \pi_1(\mathbf{Y},\mathbf{y}_1)$, then define $\tilde{\mu} : [0,1] \to \mathbf{Y}$ by

$$\begin{aligned}
\tilde{\mu}(t) &= \tilde{\alpha}(3t), & t &\in \left[0,\frac{1}{3}\right],\\
&= \tilde{\gamma}\left(\frac{3t-1}{2}\right), & t &\in \left[\frac{1}{3},\frac{2}{3}\right],\\
&= \tilde{\alpha}(3-3t), & t &\in \left[\frac{2}{3},1\right].
\end{aligned}$$

Now, set $\gamma = p \circ \tilde{\gamma}$ and $\mu = p \circ \tilde{\mu}$. It is easy to show that $[\tilde{\mu}] \in \pi_1(\mathbf{Y},\mathbf{y}_0)$ and $[\mu] = [\alpha]^{-1}[\gamma][\alpha] \in \pi_1(\mathbf{X},\mathbf{x}_0)$. See Figure 7.28.

Next, we would like to classify covering spaces. Let (**Y**,p) be a covering space for a space **X** and let $\mathbf{x}_0 \in \mathbf{X}$ and $\mathbf{y}_0 \in p^{-1}(\mathbf{x}_0)$. First, we shall answer the question about when maps from some arbitrary space **Z** into **X** lifts to a map into **Y**. Let $\mathbf{z}_0 \in \mathbf{Z}$. The specific question is, given a map $f : (\mathbf{Z},\mathbf{z}_0) \to (\mathbf{X},\mathbf{x}_0)$, when does a lifting $\tilde{f} : (\mathbf{Z},\mathbf{z}_0) \to (\mathbf{Y},\mathbf{y}_0)$ exist? In terms of diagrams, we are given f and p and are looking for an $\tilde{f}$ that will produce a commutative diagram

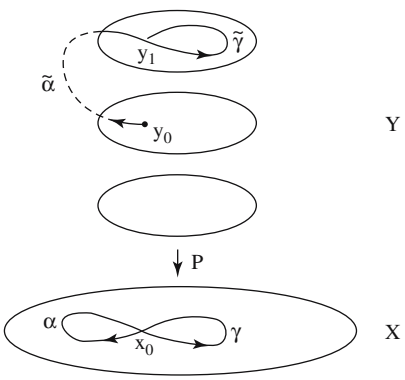



**Figure 7.28.** How loops in the total space project to conjugate loops.

$$\begin{tikzcd} & \mathbf{Y},\mathbf{y}_0 \arrow[d, "p"] \\ \mathbf{Z},\mathbf{z}_0 \arrow[ur, "\tilde{f}"] \arrow[r, "f"] & \mathbf{X},\mathbf{x}_0. \end{tikzcd}$$

A necessary condition is clearly that $f_*(\pi_1(\mathbf{Z},\mathbf{z}_0)) \subseteq p_*(\pi_1(\mathbf{Y},\mathbf{y}_0))$. The amazing fact is that this condition is also sufficient provided that some weak connectivity conditions hold.

**Definition.** A topological space $\mathbf{X}$ is said to be *locally path-connected* if every neighborhood of a point contains a neighborhood that is path-connected.

Fortunately, the spaces of interest to us are locally path-connected. Manifolds are trivially locally path-connected, but so are CW complexes (use induction on the number of cells).

**7.4.2.11. Theorem.** Let $(\mathbf{Y},p)$ be a covering space for a space $\mathbf{X}$. Let $\mathbf{Z}$ be a path-connected and locally path-connected space. Let $\mathbf{x}_0 \in \mathbf{X}$, $\mathbf{y}_0 \in p^{-1}(\mathbf{x}_0)$, and $\mathbf{z}_0 \in \mathbf{Z}$. Then a map $f : (\mathbf{Z},\mathbf{z}_0) \to (\mathbf{X},\mathbf{x}_0)$ lifts to a map $\tilde{f} : (\mathbf{Z},\mathbf{z}_0) \to (\mathbf{Y},\mathbf{y}_0)$ if and only if $f_*(\pi_1(\mathbf{Z},\mathbf{z}_0)) \subseteq p_*(\pi_1(\mathbf{Y},\mathbf{y}_0))$.

**Proof.** See [Mass67] or [Jäni84]. The diagram below should help clarify what is being said:

$$\begin{tikzcd} & \mathbf{Y},\mathbf{y}_0 \arrow[d, "p"] & \pi_1(\mathbf{Y},\mathbf{y}_0) \arrow[d, "p_*"] \\ \mathbf{Z},\mathbf{z}_0 \arrow[ur, "\tilde{f}"] \arrow[r, "f"] & \mathbf{X},\mathbf{x}_0 & p_*(\pi_1(\mathbf{Y},\mathbf{y}_0)) \\ & & \rotatebox{90}{$\subseteq$} \\ \pi_1(\mathbf{Z},\mathbf{z}_0) \arrow[rr, "f_*"] & & f_*(\pi_1(\mathbf{Z},\mathbf{z}_0)) \subseteq \pi_1(\mathbf{X},\mathbf{x}_0). \end{tikzcd}$$

We can deduce a number of important results from Theorem 7.4.2.11.

**7.4.2.12. Theorem.** Let $(\mathbf{Y}_1,p_1)$ and $(\mathbf{Y}_2,p_2)$ be covering spaces for a space $\mathbf{X}$, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are path-connected and locally path-connected spaces. Let $\mathbf{x}_0 \in \mathbf{X}$ and $\mathbf{y}_i \in p_i^{-1}(\mathbf{x}_0)$. The two covering spaces are isomorphic via a bundle isomorphism $f : (\mathbf{Y}_1,\mathbf{y}_1) \to (\mathbf{Y}_2,\mathbf{y}_2)$ if and only if $p_{1*}(\pi_1(\mathbf{Y}_1,\mathbf{y}_1)) = p_{2*}(\pi_1(\mathbf{Y}_2,\mathbf{y}_2))$.

**Proof.** See [Mass67]. The following diagram might again help:

$$\begin{tikzcd} \pi_1(\mathbf{Y}_1,\mathbf{y}_1) \arrow[d, "p_{1*}"'] & \mathbf{Y}_1 \arrow[r, "f"] \arrow[dr, "p_1"'] & \mathbf{Y}_2 \arrow[dl, "p_2"] & \pi_1(\mathbf{Y}_2,\mathbf{y}_2) \arrow[d, "p_{2*}"] \\ & & \mathbf{X} & \end{tikzcd}$$

$$p_{1*}(\pi_1(\mathbf{Y}_1,\mathbf{y}_1)) \subseteq \pi_1(\mathbf{X},\mathbf{x}_0) \supseteq p_{2*}(\pi_1(\mathbf{Y}_2,\mathbf{y}_2)).$$

To get the next theorem we need another technical definition.

**Definition.** A topological space $\mathbf{X}$ is said to be *semi-locally simply connected* if every point $\mathbf{x}$ in $\mathbf{X}$ has a neighborhood $\mathbf{U}$ so that every closed curve in $\mathbf{U}$ that starts at $\mathbf{x}$ is homotopic to a constant map in $\mathbf{X}$.

Manifolds and CW complexes are semi-locally simply connected (use induction on the number of cells for CW complexes).

**7.4.2.13. Theorem.** Let $\mathbf{X}$ be a path-connected, locally path-connected, and semi-locally simply connected space. Let $\mathbf{x}_0 \in \mathbf{X}$. If G is an arbitrary subgroup of $\pi_1(\mathbf{X},\mathbf{x}_0)$, then there is a path-connected and locally path-connected space $\mathbf{Y}$ and covering space $(\mathbf{Y},\mathrm{p})$ of $\mathbf{X}$, so that for some point $\mathbf{y}_0 \in \mathrm{p}^{-1}(\mathbf{x}_0)$, $\mathrm{p}_*(\pi_1(\mathbf{Y},\mathbf{y}_0)) = \mathrm{G}$.

**Proof.** See [Mass67] or [Jäni84].

**Definition.** A *universal cover* or *universal covering space* for a space $\mathbf{X}$ is a covering space $(\mathbf{Y},\mathrm{p})$ for $\mathbf{X}$, with the property that $\mathbf{Y}$ is path-connected, locally path-connected, and simply connected.

By Theorem 7.4.2.12, the universal covering space of a space (if it exists) is unique up to isomorphism. Therefore, if the projection p is obvious from the context, then the common expression "the universal cover $\mathbf{Y}$ of $\mathbf{X}$" refers to the universal covering space $(\mathbf{Y},\mathrm{p})$.

**7.4.2.14. Example.** The space $\mathbf{R}$ is the universal cover of the circle $\mathbf{S}^1$ (see Example 7.4.2.2).

**7.4.2.15. Example.** The sphere $\mathbf{S}^n$ is the universal cover of projective space $\mathbf{P}^n$ (see Example 7.4.2.1).

**7.4.2.16. Theorem.** Let $\mathbf{X}$ be a path-connected, locally path-connected and semi-locally simply connected space. Then $\mathbf{X}$ has a universal covering space and any two are isomorphic.

**Proof.** Only the existence part of this theorem needs proving. See [Mass67] or [Jäni84].

The reason that a universal covering space $(\mathbf{Y},\mathrm{p})$ for a space $\mathbf{X}$ has the name it has is that if $(\mathbf{Y}',\mathrm{p}')$ is any other covering space for $\mathbf{X}$, then there a unique (up to isomorphism) map $\tilde{\mathrm{p}} : \mathbf{Y} \to \mathbf{Y}'$ making the following diagram commutative

$$\begin{tikzcd} \mathbf{Y} \arrow[rr, "\tilde{\mathrm{p}}"] \arrow[dr, "\mathrm{p}"'] & & \mathbf{Y}' \arrow[dl, "\mathrm{p}'"] \\ & \mathbf{X} & \end{tikzcd}.$$

In fact, $(\mathbf{Y},\tilde{\mathrm{p}})$ will be a covering space for $\mathbf{Y}'$. In other words, the universal covering space of a space "covers" every other covering space of the space.

The covering transformations of a covering space are interesting. They obviously form a group.

**Definition.** Cov($\mathbf{Y}$,p) will denote the group of covering transformations of a covering space ($\mathbf{Y}$,p).

**7.4.2.17. Theorem.** Let ($\mathbf{Y}$,p) be a covering space for a path-connected and locally path-connected space $\mathbf{X}$. Let $\mathbf{x}_0 \in \mathbf{X}$ and $\mathbf{y}_0 \in p^{-1}(\mathbf{x}_0)$. Let $G = p_*(\pi_1(\mathbf{Y},\mathbf{y}_0))$ and let $N_G$ be the normalizer of G in $\pi_1(\mathbf{X},\mathbf{x}_0)$. Given an element $[\gamma] \in N_G$, there is exactly one covering transformation $h_{[\gamma]}$ that maps $\mathbf{y}_0$ into the end point $\tilde{\gamma}(1)$ of the lifting $\tilde{\gamma}$ of $\gamma$ that starts at $\mathbf{y}_0$. The map

$$\begin{aligned} N_G &\to \mathrm{Cov}(\mathbf{Y},p) \\ [\gamma] &\to h_{[\gamma]} \end{aligned}$$

is a homomorphism with kernel G, that is,

$$\mathrm{Cov}(\mathbf{Y},p) \approx \frac{N_G}{G}.$$

**Proof.** See [Mass67] or [Jäni84].

**7.4.2.18. Corollary.** Let ($\mathbf{Y}$,p) be the universal covering space for a path-connected and locally path-connected space $\mathbf{X}$ and let $\mathbf{x}_0 \in \mathbf{X}$. Then $\mathrm{Cov}(\mathbf{Y},p) \approx \pi_1(\mathbf{X},\mathbf{x}_0)$. If $\pi_1(\mathbf{X},\mathbf{x}_0)$ is finite and $n = |\pi_1(\mathbf{X},\mathbf{x}_0)|$, then ($\mathbf{Y}$,p) is an n-fold covering.

**7.4.2.19. Example.** The covering transformations of the universal covering space ($\mathbf{R}$,p) defined in Example 7.4.2.2 are the maps

$$h_n : \mathbf{R} \to \mathbf{R}$$

defined by

$$h_n(t) = t + 2\pi n.$$

The maps $h_n$ are obviously covering transformations. Note that $h_n = h_1^n$.

**7.4.2.20. Example.** The only covering transformation of the covering space $\mathbf{S}^n$ over $\mathbf{P}^n$ is the antipodal map of $\mathbf{S}^n$.

**7.4.2.21. Corollary.** Let ($\mathbf{Y}$,p) be the universal covering space for a path-connected and locally path-connected space $\mathbf{X}$ and let $\mathbf{x}_0 \in \mathbf{X}$. Then $\mathbf{X} \approx \mathbf{Y}/\sim$, where ~ is the equivalence relation defined by $\mathbf{y} \sim \mathbf{y}'$ if there is an $h \in \mathrm{Cov}(\mathbf{Y},p)$, such that $\mathbf{y}' = h(\mathbf{y})$.

**7.4.2.22. Example.** Let $m, n \in \mathbf{Z}$. Define maps

$$h_{m,n} : \mathbf{R}^2 \to \mathbf{R}^2$$

by

$$h_{m,n}(x,y) = (x + 2\pi m, y + 2\pi n)$$

and think of the torus $\mathbf{S}^1 \times \mathbf{S}^1$ as the quotient space of $\mathbf{R}^2/\sim$, where $\sim$ is the equivalence relation

$$(x,y) \sim h_{m,n}(x,y).$$

If $p : \mathbf{R}^2 \to \mathbf{R}^2/\sim$ is the quotient map, then $(\mathbf{R}^2,p)$ is the universal covering space of the torus. With this interpretation, the maps $h_{m,n}$ are obviously the covering transformations. Note that $h_{m,n} = h^m_{1,0} \circ h^n_{0,1}$.

Using Corollary 7.4.2.18 and what we showed in Examples 7.4.2.19, 7.4.2.20, and 7.4.2.22 we now have alternate proofs of the facts stated in Theorem 7.4.1.10(1) and (2) and Corollary 7.4.1.15, namely,

**7.4.2.23. Corollary.**

(1) $\pi_1(\mathbf{S}^1) \approx \mathbf{Z}$.
(2) $\pi_1(\mathbf{P}^n) \approx \mathbf{Z}_2$.
(3) $\pi_1(\mathbf{S}^1 \times \mathbf{S}^1) \approx \mathbf{Z} \oplus \mathbf{Z}$.

All this talk about covering transformations and the last three examples leads to another question. Suppose that we turn things around and start with a group of homeomorphisms G of a space $\mathbf{Y}$ and define

$$\mathbf{Y}/\mathbf{G} = \mathbf{Y}/\sim, \tag{7.8}$$

where

$$\mathbf{y}_1 \sim \mathbf{y}_2 \quad \text{if} \quad \mathbf{y}_2 = h(\mathbf{y}_1) \quad \text{for some } h \in G.$$

**Definition.** The space $\mathbf{Y}/G$ in equation (7.8) is called the *quotient space* of $\mathbf{Y}$ *modulo the group* G.

If $p : \mathbf{Y} \to \mathbf{Y}/G$ is the quotient map, then is $(\mathbf{Y},p,\mathbf{Y}/G)$ a covering space with G the group of covering transformations? The answer in general is no. At the very least the homeomorphisms in G could not be allowed to have fixed points, but we need something stronger.

**Definition.** A group of homeomorphism G of a space $\mathbf{X}$ is said to be *properly discontinuous* if every point $\mathbf{x}$ in $\mathbf{X}$ has a neighborhood $\mathbf{U}$ so that all the sets $h(\mathbf{U})$, $h \in G$, are disjoint.

Clearly, no homeomorphism in a properly discontinuous group of homeomorphism can have a fixed point. Furthermore, it is easy to see that the covering transformations of a covering space form a properly discontinous group of homeomorphisms of the total space.

**7.4.2.24. Theorem.** Let $\mathbf{Y}$ be a connected, locally path-connected topological space and G a properly discontinuous group of homeomorphism of $\mathbf{Y}$. If $p : \mathbf{Y} \to \mathbf{Y}/G$ is the

quotient map, then (**Y**p) is a covering space for **Y**/G with G the group of covering transformations.

**Proof.** See [Mass67].

As a nice application of this discussion of covering transformations we relate this to the lens spaces defined in Section 7.2.4. As before, let p and q be relatively prime positive integers and assume that $0 \le q \le p/2$. Consider the unit sphere $\mathbf{S}^3$ in complex 2-space $\mathbf{C}^2$. (One can identify $\mathbf{C}^2$ with $\mathbf{R}^4$.) Define rotations

$$\sigma_i : \mathbf{S}^3 \to \mathbf{S}^3$$

by

$$\sigma_i(\mathbf{z}_1, \mathbf{z}_2) = \left(\mathbf{z}_1 e^{2\pi i/p}, \mathbf{z}_2 e^{2\pi q i/p}\right),$$

where $\mathbf{z}_1$, $\mathbf{z}_2 \in \mathbf{C}$ and i = 0, 1, . . . , p–1. (The $e^{2\pi i/p}$ are the pth roots of unity.) Since $\sigma_i = \sigma_0^i$, we have in effect defined a group

$$G = \{\sigma_0, \sigma_1, \ldots, \sigma_{p-1}\}$$

of order p of rotations acting on $\mathbf{S}^3$. Define

$$L(p,q) = \mathbf{S}^3/G$$

and let

$$\eta : \mathbf{S}^3 \to L(p,q)$$

be the quotient map.

**7.4.2.25. Theorem.** The new spaces L(p,q) are homeomorphic to the lens spaces L(p,q) defined in Section 7.2.4. Furthermore,

(1) $(\mathbf{S}^3, \eta)$ is the universal covering space for L(p,q).
(2) $G = \mathrm{Cov}(\mathbf{S}^3, \eta)$.

**Proof.** See [CooF67] or [HilW60]. Parts (1) and (2) are obvious.

**7.4.2.26. Corollary.** $\pi_1(L(p,q)) \approx \mathbf{Z}_p$.

This concludes our overview of the theory of covering spaces. The gist of the main results stripped of their technical details is summarized by the following:

(1) The theory of covering spaces for simply connected spaces is uninteresting because the only covering space in that case is where the space covers itself.
(2) The universal covering space of a space covers every other covering space of the space.

(3) The covering transformations of the universal covering space of a space have no fixed points and are in one-to-one correspondence with the elements of the fundamental group of the space. They act transitively on the fibers, that is, any point in the total space can be mapped into any other point belonging to the same fiber as the first.

(4) Every conjugacy class of a subgroup of the fundamental group of a space defines a covering space for that space.

See [Jäni84] for a variety of applications of covering space theory.

### 7.4.3 Higher Homotopy Groups

There is an important generalization of the fundamental group of a space that leads to higher-dimensional homotopy groups.

**Definition.** Let $n \geq 2$. Given maps $\alpha, \beta : (\mathbf{I}^n, \partial\mathbf{I}^n) \to (\mathbf{X}, \mathbf{x}_0)$, define a map

$$\alpha * \beta : \left(\mathbf{I}^n, \partial\mathbf{I}^n\right) \to (\mathbf{X}, \mathbf{x}_0)$$

by

$$\begin{aligned}(\alpha * \beta)(t_1, t_2, \ldots, t_n) &= \alpha(2t_1, t_2, \ldots, t_n), \qquad \text{if } 0 \leq t_1 \leq \frac{1}{2}, \\ &= \beta(2t_1 - 1, t_2, \ldots, t_n), \quad \text{if } \frac{1}{2} \leq t_1 \leq 1.\end{aligned}$$

**Definition.** Let

$$\pi_n(\mathbf{X}, \mathbf{x}_0) = \left\{\alpha \mid \alpha : \left(\mathbf{I}^n, \partial\mathbf{I}^n\right) \to (\mathbf{X}, \mathbf{x}_0)\right\} \big/ \simeq_{\partial\mathbf{I}^n}$$

be the set of equivalence classes of maps $\alpha$ with respect to the equivalence relation $\simeq_{\partial\mathbf{I}^n}$. Define a product $*$ on $\pi_n(\mathbf{X}, \mathbf{x}_0)$ as follows: If $[\alpha], [\beta] \in \pi_n(\mathbf{X}, \mathbf{x}_0)$, then

$$[\alpha] * [\beta] = [\alpha * \beta].$$

**7.4.3.1. Theorem.** The operation $*$ on $\pi_n(\mathbf{X}, \mathbf{x}_0)$ is well defined and makes $\pi_n(\mathbf{X}, \mathbf{x}_0)$ into a group called the *nth homotopy group* of the pointed space $(\mathbf{X}, \mathbf{x}_0)$.

**Proof.** The proof is similar to the one for the fundamental group. Exercise 7.4.3.1.

There is a perhaps easier way to visualize the product in $\pi_n(\mathbf{X}, \mathbf{x}_0)$. First, we need a definition.

**Definition.** Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pointed spaces with base points $\mathbf{x}_0$, $\mathbf{y}_0$, and $\mathbf{z}_0$, respectively. If $f : \mathbf{X} \to \mathbf{Z}$ and $g : \mathbf{Y} \to \mathbf{Z}$ are continuous maps with $f(\mathbf{x}_0) = \mathbf{z}_0$ and $g(\mathbf{y}_0) = \mathbf{z}_0$, then define a map

$$f \vee g : \mathbf{X} \vee \mathbf{Y} \to \mathbf{Z} \quad \text{by} \quad f \vee g \,|\, \mathbf{X} = f \quad \text{and} \quad f \vee g \,|\, \mathbf{Y} = g.$$

This map is called the *wedge* of f and g.

With this definition we can now give an alternate definition of the product in $\pi_n(\mathbf{X},\mathbf{x}_0)$. First of all, we can clearly identify maps $(\mathbf{I}^n,\partial\mathbf{I}^n) \to (\mathbf{X},\mathbf{x}_0)$ with maps $(\mathbf{S}^n,\mathbf{e}_1) \to (\mathbf{X},\mathbf{x}_0)$. Let

$$c : \mathbf{S}^n \to \mathbf{S}^n \vee \mathbf{S}^n \tag{7.9}$$

be the map that collapses $\mathbf{S}^{n-1}$ to the base point of $\mathbf{S}^n \vee \mathbf{S}^n$ and that wraps the upper and lower hemisphere of $\mathbf{S}^n$ around the first and second factor of $\mathbf{S}^n \vee \mathbf{S}^n$, respectively. Let $[\alpha]$, $[\beta] \in \pi_n(\mathbf{X},\mathbf{x}_0)$. If we represent $\alpha$ and $\beta$ as maps

$$\alpha,\beta : (\mathbf{S}^n,\mathbf{e}_1) \to (\mathbf{X},\mathbf{x}_0),$$

then the product $[\alpha]*[\beta]$ is nothing but the homotopy class of the composite map $(\alpha\vee\beta) \circ c$. See Figure 7.29.

We now have homotopy groups $\pi_n(\mathbf{X},\mathbf{x}_0)$ defined for $n \geq 1$. It is convenient to make a definition for $n = 0$. Note that $\mathbf{S}^0 = \{-1,+1\}$.

**Definition.** $\pi_0(\mathbf{X},\mathbf{x}_0)$ is defined to be the **set** (there is no group structure) of homotopy classes of maps

$$f : (\mathbf{S}^0,1) \to (\mathbf{X},\mathbf{x}_0).$$

(Equivalently, $\pi_0(\mathbf{X},\mathbf{x}_0)$ is the set of path components of $\mathbf{X}$.)

Although $\pi_0(\mathbf{X},\mathbf{x}_0)$ has no group structure, one often refers to it as the 0th homotopy "group."

**7.4.3.2. Theorem.** The group $\pi_n(\mathbf{X},\mathbf{x}_0)$ is abelian whenever $n \geq 2$.

**Proof.** Figure 7.30 shows how to construct a homotopy between $\alpha*\beta$ and $\beta*\alpha$.

Just as in the case of the fundamental group, higher homotopy groups are independent of the base point if the space is path-connected. One therefore often writes $\pi_n(\mathbf{X})$ instead of $\pi_n(\mathbf{X},\mathbf{x}_0)$.

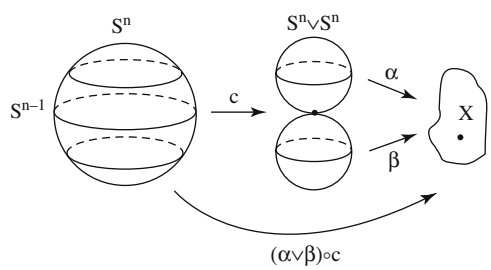



**Figure 7.29.** Using the wedge operation to define a homotopy group product.

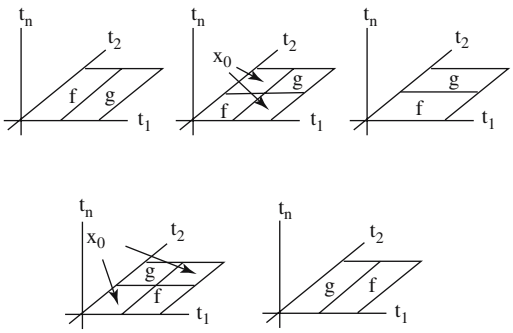



**Figure 7.30.** Proving the commutativity of homotopy groups.

Continuous maps induce homomorphisms on the homotopy groups in a natural way, similar to how it was done in the case of the fundamental group. Let $(\mathbf{X},\mathbf{x}_0)$ and $(\mathbf{Y},\mathbf{y}_0)$ be pointed spaces. Given a continuous map

$$f:\ (\mathbf{X},\mathbf{x}_0) \to (\mathbf{Y},\mathbf{y}_0)$$

define

$$f_*:\ \pi_n(\mathbf{X},\mathbf{x}_0) \to \pi_n(\mathbf{Y},\mathbf{y}_0), \quad n \geq 0,$$

by

$$f_*([\alpha]) = [f \circ \alpha].$$

**7.4.3.3. Lemma.** The map $f_*$ is well defined. It is a homomorphism of groups when $n \geq 1$.

**Proof.** Exercise 7.4.3.2.

**Definition.** The map $f_*$ is called the homomorphism *induced* by the continuous map f.

The fact that the higher ($n \geq 2$) homotopy groups are abelian sets them apart from the fundamental group. In other ways, they satisfy similar properties however. For example, one can show, just like in Theorem 7.4.1.14, that there are isomorphisms

$$\pi_n(\mathbf{X} \times \mathbf{Y}, \mathbf{x}_0 \times \mathbf{y}_0) \to \pi_n(\mathbf{X},\mathbf{x}_0) \times \pi_n(\mathbf{Y},\mathbf{y}_0).$$

(By the way, no such isomorphism exists for homology. A theorem, the Künneth theorem, relates the homology of the product of two spaces to that of the spaces but it is much more complicated.) There are also natural homomorphisms

$$\mu:\ \pi_n(\mathbf{X},\mathbf{x}_0) \to H_n(\mathbf{X}), \tag{7.10}$$

called the *Hurewicz homomorphisms*, which generalize the homomorphism for the fundamental group. (We are again pretending that $H_n(\mathbf{X})$ is well defined.) These homomorphisms are neither onto nor one-to-one in general though. The homotopy groups capture the idea of "holes" better than the homology groups. After all, the n-sphere $\mathbf{S}^n$ is the prototype of an n-dimensional "hole."

Nontrivial "higher" homotopy groups of spheres are one important example of what sets homotopy groups apart from homology groups. The homology groups $H_i(\mathbf{X})$ are all $\mathbf{0}$ if i is larger than the dimension of $\mathbf{X}$, but this is not necessarily the case for homotopy groups. For example,

$$\pi_3(\mathbf{S}^2) \approx \mathbf{Z}.$$

One well-known theorem that relates the homotopy and homology groups in a special case is

**7.4.3.4. Theorem.** (The Hurewicz Isomorphism Theorem) If $n \geq 2$ and if $\mathbf{X}$ is a connected polyhedron whose first $n - 1$ homotopy groups vanish, then the Hurewicz homorphism

$$\mu : \pi_n(\mathbf{X}, \mathbf{x}_0) \to H_n(\mathbf{X})$$

is an isomorphism.

**Proof.** See [Span66].

Theorem 7.4.3.4 is one result that can be used to compute higher homotopy groups.

**7.4.3.5. Theorem.** Let $n \geq 1$.

(1) $\pi_i(\mathbf{S}^n) = 0$ for $0 \leq i < n$.
(2) $\pi_n(\mathbf{S}^n) \approx \mathbf{Z}$.

**Proof.** To prove (1) consider a map $f : \mathbf{S}^i \to \mathbf{S}^n$. The map f is homotopic to a map that misses a point, say $\mathbf{e}_{n+1}$. (To prove this fact, use the simplicial approximation theorem with respect to some triangulations of the spheres.) But $\mathbf{S}^n - \mathbf{e}_{n+1}$ is homeorphic to an open disk that is contractible. It follows that f is homotopic to a constant map and proves (1). The case $n = 1$ in part (2) is just Corollary 7.4.2.23(1). If $n > 1$, then (2) follows from (1), Theorem 7.4.3.4, and Theorem 7.2.3.4.

It should be noted that Theorem 7.4.3.4 implies nothing about the homomorphisms

$$\mu : \pi_i(\mathbf{X}, \mathbf{x}_0) \to H_i(\mathbf{X})$$

for $i > n$ (n as in the theorem). In general, homotopy groups are much harder to compute than homology groups but they are stronger invariants than homology groups. As an example of the latter, there is the following theorem:

**7.4.3.6. Theorem.** A continuous map $f : \mathbf{X} \to \mathbf{Y}$ between CW complexes is a homotopy equivalence if and only if it induces isomorphisms on all ($n \geq 0$) homotopy groups.

**Proof.** See [LunW69].

The best that one can do for homology is

**7.4.3.7. Theorem.** Let $\mathbf{X}$ and $\mathbf{Y}$ be simply connected CW complexes. A continuous map $f : \mathbf{X} \to \mathbf{Y}$ is a homotopy equivalence if and only it induces isomorphisms on all homology groups.

**Proof.** See [Span66].

Theorem 7.4.3.7 is false if $\mathbf{X}$ and $\mathbf{Y}$ are not simply connected. For a counterexample, see [Span66], page 420. There is an analog of the theorem when spaces are not simply connected, but things get much more involved. It is based on the notion of a *simple homotopy equivalence*. See [Dieu89].

We finish with one last application.

**Definition.** A topological space that has the homotopy type of the n-sphere $\mathbf{S}^n$ is called a *homotopy n-sphere*. A polyhedron is called a *homology n-sphere* if it has the same homology groups as $\mathbf{S}^n$.

**7.4.3.8. Theorem.** Every simply connected homology n-sphere, $n \geq 2$, is a homotopy n-sphere.

**Proof.** By Theorems 7.4.3.4 we get a map between the $\mathbf{S}^n$ and the space that is an isomorphism on homology. Now use Theorem 7.4.3.7.

Theorem 7.4.3.8 is also false if we drop the simply connected hypothesis. In [SeiT80] one can find an example of a three-dimensional space, called a *Poincaré space*, that is a homology 3-sphere but that has a nontrivial fundamental group and hence cannot be of the same homotopy type as $\mathbf{S}^3$ (nor homeomorphic to it). The nice thing about Theorems 7.4.3.7 and 7.4.3.8 is that, in order to prove something about homotopy type, we do not have to mess around with complicated homotopy groups but can simply work with homology groups, which is a much easier task. We do have to check that spaces are simply connected though.

## 7.5 Pseudomanifolds

This section specializes to manifold-like spaces. We shall define what it means for them to be orientable and relate this concept to homology. Some applications of this can be found in the next section.

**Definition.** A polyhedron $\mathbf{X}$ is called an *n-dimensional pseudomanifold* or *n-pseudomanifold* or simply *pseudomanifold* if it admits a triangulation $(K,\varphi)$ satisfying

(1) Every simplex of K is a face of some n-simplex in K.
(2) Every (n – 1)-simplex of K is a face of at least one, but not more than two, n-simplices of K.
(3) Given any two n-simplices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in K there is a chain $\boldsymbol{\alpha} = \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_k = \boldsymbol{\beta}$ of n-simplices $\boldsymbol{\sigma}_i$ in K so that $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_{i+1}$ meet in an (n – 1)-simplex.

The pseudomanifold is said to be *closed* if $\partial K = \boldsymbol{\phi}$.

Note that condition (3) in the definition implies that pseudomanifolds are compact connected spaces. This is not an essential but convenient standard assumption. It is easy to see that every (combinatorial) surface is a pseudomanifold, but not every two-dimensional pseudomanifold is a surface. Figure 7.31(a) shows a two-dimensional pseudomanifold with boundary that is not a surface with boundary. Figure 7.31(b) shows a *pinched sphere* (a sphere with two points identified). The problem occurs at the points **p** that do not have the correct neighborhood. In general, every triangulable n-manifold is an n-pseudomanifold. Although the boundary of every manifold is a manifold, Figure 7.31(a) also shows that this need not be the case for pseudomanifolds. Nevertheless, pseudomanifolds have enough nice manifold-type properties, so that they are interesting because many properties of manifolds are true simply because they satisfy the pseudomanifold conditions. Finally, one nice fact (Theorem 7.5.2) is that it does not matter how we triangulate a pseudomanifold because every triangulation will satisfy properties (1)–(3). One way to prove this topological invariance of the combinatorial structure of a pseudomanifold is to establish the following interesting property of the top-dimensional mod 2 homology group of a pseudomanifold (and hence manifold) first.

**7.5.1. Theorem.** Let **X** be an n-dimensional pseudomanifold.

(1) If $\partial\mathbf{X} = \boldsymbol{\phi}$, then $H_n(\mathbf{X};\mathbf{Z}_2) \approx \mathbf{Z}_2$.
(2) If $\partial\mathbf{X} \neq \boldsymbol{\phi}$, then $H_n(\mathbf{X};\mathbf{Z}_2) = 0$.

**Proof.** See [AgoM76]. The proofs are not hard. They are similar to our computations of homology groups for $\mathbf{S}^n$ and consists in finding the obvious cycles and boundaries. Note that over $\mathbf{Z}_2$ the orientation of simplices does not play a role.

**7.5.2. Theorem.** (Invariance of Pseudomanifolds) Let **X** be an n-dimensional pseudomanifold and let $(L,\Psi)$ be any triangulation of **X**. Then L satisfies properties (1)–(3) in the definition of a pseudomanifold.


**Figure 7.31.** Pseudomanifolds that are not manifolds.

**Proof.** See [AgoM76]. The proof is also not hard but too long to give here. It does make use of Theorem 7.5.1.

To define the orientability of pseudomanifolds, we start with a combinatorial definition and then show how one can use a homology group to detect this property.

**Definition.** Let **X** be an n-dimensional pseudomanifold and let $(K,\varphi)$ be a triangulation of **X**. We shall say that **X** is *orientable* if the n-simplices of K can be oriented coherently, that is, the n-simplices of K can be oriented simultaneously in such a way that any two n-simplices that meet in a common $(n-1)$-dimensional face induce opposite orientations on that face. Such a choice of orientations of n-simplices, if it exists, is called an *orientation* of **X**. If **X** is not orientable, one calls **X** *nonorientable*.

The next theorem shows that the definition is well defined and independent of the particular triangulation that is chosen. It also shows that orientability is easily determined.

**7.5.3. Theorem.** A closed n-dimensional pseudomanifold **X** is orientable if and only if $H_n(\mathbf{X}) \approx \mathbf{Z}$.

**Proof.** See [AgoM76]. The proof again relies on finding the right cycles like in the past.

Using the results in Table 7.2.1.1 for surfaces, we see that our new rigorous definition of orientable agrees with our previous intuitive definition. More importantly, we now have an algorithm for determining the orientability of a surface. It is also clear that choosing an orientation of a closed n-dimensional pseudomanifold **X** is equivalent to choosing a generator of $H_n(\mathbf{X})$. Theorem 7.5.1 shows that the mod 2 homology groups tell us nothing about the orientability of **X**.

Before we state another useful criterion for when a pseudomanifold is orientable we need to discuss a few more concepts associated to pseudomanifolds.

Let K be a simplicial complex that triangulates an n-dimensional pseudomanifold $\mathbf{M}^n$. Recall the definition of the barycentric subdivision sd(K) of the simplicial complex K given in Section 7.2.2. Its vertices are the barycenters of the simplices in K. If $b(\boldsymbol{\sigma})$ again denotes the barycenter of the simplex $\boldsymbol{\sigma}$, then the k-simplices of sd(K), $k > 0$, are all the k-simplices of the form $b(\boldsymbol{\sigma}_0)b(\boldsymbol{\sigma}_1)\ldots b(\boldsymbol{\sigma}_k)$ where the $\boldsymbol{\sigma}_i$ are distinct simplices of K and $\boldsymbol{\sigma}_0 \prec \boldsymbol{\sigma}_1 \prec \ldots \prec \boldsymbol{\sigma}_k$.

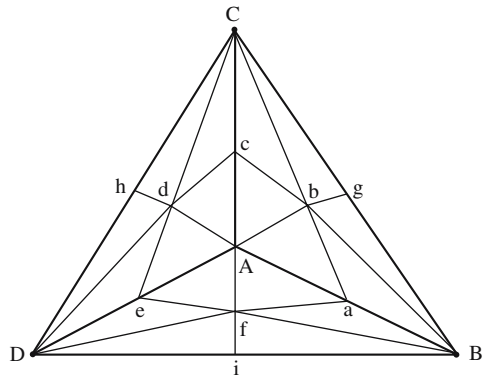
**Definition.** Let $\boldsymbol{\sigma}^k$ be a k-simplex of K. Define the *dual* $(n-k)$*-cell* $\boldsymbol{\sigma}_*^{n-k}$ by

$$\boldsymbol{\sigma}_*^{n-k} = \cup\,\{b(\boldsymbol{\sigma})b(\boldsymbol{\sigma}_1)b(\boldsymbol{\sigma}_2)\ldots b(\boldsymbol{\sigma}_{n-k}) \mid \boldsymbol{\sigma}_i \text{ is a simplex in K and } \boldsymbol{\sigma} \prec \boldsymbol{\sigma}_1 \prec \boldsymbol{\sigma}_2 \prec \ldots \prec \boldsymbol{\sigma}_{n-k}\}.$$

Call $b(\boldsymbol{\sigma})$ the *barycenter* of $\boldsymbol{\sigma}_*^{n-k}$.

**7.5.4. Example.** Consider the two-dimensional simplicial complex K shown in Figure 7.32 whose vertices are labeled with uppercase letters and whose edges are drawn with thick lines. The barycentric subdivision of K is drawn with thin lines and its additional vertices are labeled with lowercase letter. The dual cell of the 0-simplex **A** is the union of the 2-cells **Aab**, **Abc**, **Acd**, **Ade**, **Aef**, and **Afa** in the barycentric

**Figure 7.32.** Defining the dual cells.




subdivision. The dual cell of the 1-simplex **AB** is the union of the two 1-simplices **ab** and **af**. The dual of the 2-simplex **ABC** is the 0-simplex **b**.

The dual cell $\boldsymbol{\sigma}_*^{n-k}$ will not necessarily be homeomorphic to $\mathbf{D}^{n-k}$. It is if we are dealing with a proper triangulation for a combinatorial manifold but not for an arbitrary pseudomanifold. For example, the dual cell of the vertices **p** in Figure 7.31 are not disks. In order for the dual cells to have nice properties we do not need anything as strong as a combinatorial manifold.

**Definition.** A closed compact connected topological n-dimensional manifold is called a *homology manifold* if it admits a triangulation (K,φ) with the property that the boundary of the star of every vertex in K is a homology (n – 1)-sphere.

Earlier results in this chapter show that the property of being a homology manifold is a topological invariant.

**7.5.5. Theorem.**

(1) Every homology manifold is a pseudomanifold.
(2) An n-dimensional homology manifold is a manifold when $n \le 3$.

**Proof.** See [SeiT80] or [Cair68].

Not every n-dimensional homology manifolds is a manifold when $n > 4$ (see [SeiT80] for counterexamples), but they have enough in common to be able to prove the important duality theorems in Section 7.5.2.

Returning to the dual of a k-simplex $\boldsymbol{\sigma}^k$ in a simplicial complex K that triangulates a pseudomanifold $\mathbf{M}^n$, we have

**7.5.6. Proposition**.

(1) $\boldsymbol{\sigma}_*^{n-k}$ is an (n – k)-dimensional pseudomanifold.
(2) If **M** is a homology manifold, then $\partial\boldsymbol{\sigma}_*^{n-k}$ is an (n – k – 1)-dimensional homology sphere.
(3) The simplex $\boldsymbol{\sigma}^k$ and its dual cell $\boldsymbol{\sigma}_*^{n-k}$ intersect in the single point $b(\boldsymbol{\sigma})$.

**Proof.** See [SeiT80] or [Cair68].

**Definition.** The collection of dual cells of K, denoted by $K_*$, is called the *dual cell complex* of K.

If the dual cells were homeomorphic to disks, then we really would have a CW complex as defined in Section 7.2.4. Even without this, $K_*$ has many properties in common with such a cell complex because the dual cells are contractible and one can compute the homology groups from them. See [SeiT80] or [Cair68]. Note further that since we have a notion of a barycenter of dual cells, we can define duals of the cells in $K_*$ in the same way that we defined the duals of cells in K. With this definition of the dual cell complex of the dual complex, it is easy to see that $(K_*)_* = K$, because the dual cell of a dual cell consists of cells whose union is a cell in the original complex. We will allow ourselves to talk of the dual of a cell both in the case where it belongs to K and also when it belongs to $K_*$. We can also define what we mean by an orientation of a dual cell.

**Definition.** An *orientation* of a dual cell $\boldsymbol{\sigma}_*^{n-k}$ consists of a collection of compatible orientations of the (n – k)-simplices of $\boldsymbol{\sigma}_*^{n-k}$, that is, if two of the (n – k)-simplices meet in an (n – k – 1)-simplex $\boldsymbol{\tau}$, then they should induce opposite orientations on $\boldsymbol{\tau}$.

Let $\mathbf{M}^n$ be an n-dimensional pseudomanifold and let $(K,\varphi)$ be a triangulation of it. For simplicity, assume $\mathbf{M} = |K|$. Let $\mathbf{v}$ be a vertex of K. Consider an edge loop $\gamma = (\mathbf{v} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k = \mathbf{v})$ at $\mathbf{v}$ and the dual cells $c_i^n = (\mathbf{v}_i)_*$ and $c_i^{n-1} = (\mathbf{v}_i\mathbf{v}_{i+1})_*$ of dimension n and n – 1, respectively. Now, an orientation of $c_i^n$ induces an orientation of $c_i^{n-1}$. Conversely, it is easy to see that an orientation of $c_i^{n-1}$ induces an orientation of $c_i^n$. Therefore, if we start with an orientation o of $c_0^n$, then this will define an orientation on $c_0^{n-1}$, then on $c_1^n$, then on $c_1^{n-1}$, and so on, until we arrive back at $c_0^n$ with an induced orientation that we shall denote by $\gamma(o)$. The orientation $\gamma(o)$ may or may not agree with the original orientation o, but if we walk around the edge loop twice, that is, if we were to use the edge loop $\gamma_2 = (\mathbf{v} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k = \mathbf{v})$, then definitely $\gamma_2(o) = o$, because there are only two possible orientations for a simplex or dual cell. In any case, one can show, using the Simplicial Approximation Theorem, that, starting with an orientation o "at $\mathbf{v}$", any continuous map $\alpha : (\mathbf{I},\partial\mathbf{I}) \to (\mathbf{M},\mathbf{v})$ will induce a well-defined orientation $\alpha(o)$ at $\mathbf{v}$ that corresponds to walking around $\alpha(t)$ carrying o. It should be intuitively clear that if $\alpha(o) = o$ for all $\alpha$, then $\mathbf{M}$ is orientable. If $\alpha(o) \neq o$ for some $\alpha$, then we get a well-defined subgroup

$$H = \{[\alpha] \in \pi_1(\mathbf{M},\mathbf{v}) \mid \alpha(o) = o\}$$

of $\pi_1(\mathbf{M},\mathbf{v})$ of index 2. The subgroup H gives rise to a double covering space $(\tilde{\mathbf{M}},\tilde{\boldsymbol{\pi}})$ for $\mathbf{M}$. One can show that $\tilde{\mathbf{M}}$ is an orientable pseudomanifold. The results are summarized in the next theorem.

**7.5.7. Theorem.**

(1) Every simply connected pseudomanifold is orientable.
(2) If M is a nonorientable pseudomanifold, then there is a double covering space $(\tilde{\mathbf{M}},\tilde{\boldsymbol{\pi}})$ for $\mathbf{M}$ with $\tilde{\mathbf{M}}$ an orientable pseudomanifold and this double covering with orientable total space is unique (up to isomorphism).

**Proof.** One can find more details for the argument sketched above in [SeiT80]. Part (1) actually follows from part (2) because a simply connected space cannot have any nontrivial covering space.

### 7.5.1 The Degree of a Map and Applications

This section describes some applications related to orientable pseudomanifolds. We shall yet again assume that polyhedra have well-defined homology groups.

**Definition.** Let $\mathbf{M}^n$ be a closed orientable pseudomanifold. If $f:\mathbf{M}^n \to \mathbf{M}^n$ is a continuous map, define the *degree of f*, denoted by deg f, to be the unique integer defined by the property that

$$f_*(a) = (\deg f)\, a$$

for all $a \in H_n\mathbf{M}^n$ ($\approx \mathbf{Z}$).

**7.5.1.1. Theorem.** Let $\mathbf{M}^n$ be a closed orientable pseudomanifold and let $f, g:\mathbf{M}^n \to \mathbf{M}^n$ be continuous maps.

(1) If f is the identity map, then deg f = 1.
(2) If f is a constant map, then deg f = 0.
(3) If f and g are homotopic, then deg f = deg g.
(4) deg $(f \circ g)$ = (deg f)(deg g).
(5) If f is a homeomorphism, then deg f = ±1.

**Proof.** Easy.

To better understand the degree of a map $f:\mathbf{S}^n \to \mathbf{S}^n$ for arbitrary $n \geq 1$, note that by Theorems 7.4.1.16, 7.4.3.4, and 7.4.3.5 there is a commutative diagram

$$\begin{array}{ccc} \pi_n(\mathbf{S}^n) & \xrightarrow[f_*]{} & \pi_n(\mathbf{S}^n) \\ \mu\downarrow & & \downarrow\mu \\ H_n(\mathbf{S}^n) & \xrightarrow[f_*]{} & H_n(\mathbf{S}^n) \end{array},$$

with the Hurewicz homomorphisms $\mu$ being isomorphisms for **all** such n. Homology classes are more algebraic and less geometric than elements of homotopy groups. By looking at what happens in the diagram at the homotopy group level it is more intuitively obvious that the degree of f roughly states how many times $\mathbf{S}^n$ is wrapped around itself by f.

**7.5.1.2. Theorem.** Let $n \geq 1$. The reflection $r:\mathbf{S}^n \to \mathbf{S}^n$ defined by

$$r(x_1, x_2, \ldots, x_{n+1}) = (-x_1, x_2, \ldots, x_{n+1})$$

has degree –1.

**Proof.** It is not hard to prove the theorem by simple computations using an appropriate triangulation $(K,\varphi)$ of $\mathbf{S}^n$ and the simplicial map of K corresponding to r, but the details are rather messy and we do not repeat them here. See [AgoM76]. Intuitively, think of $\mathbf{S}^n$ as a cell complex with two n-cells that consist of two hemispheres (the parts with the x-coordinates nonnegative or nonpositive). Then a generator of $H_n(\mathbf{S}^n)$ can be represented by a cycle that consists of the sum of these two cells oriented appropriately. The reflection will then map each of these oriented n-cells into the other, but with the opposite orientation.

**7.5.1.3. Theorem.** Let $n \geq 1$. The antipodal map $f: \mathbf{S}^n \to \mathbf{S}^n$, $f(\mathbf{p}) = -\mathbf{p}$, has degree $(-1)^{n+1}$.

**Proof.** Consider the reflections $r_i: \mathbf{S}^n \to \mathbf{S}^n$, $1 \leq i \leq n+1$, defined by

$$r_i(x_1, x_2, \ldots, x_{n+1}) = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_{n+1}).$$

Clearly, $f = r_1 \circ r_2 \circ \ldots \circ r_{n+1}$. Therefore, the theorem follows easily from Theorems 7.5.1.1(4) and 7.5.1.2.

Actually, the case where n is odd can be proved directly without appealing to Theorem 7.5.1.2. In terms of coordinates,

$$f(x_1, x_2, \ldots, x_{n+1}) = (-x_1, -x_2, \ldots, -x_{n+1}).$$

If $n = 1$, then f is just a rotation through 180 degrees and is homotopic to the identity. The map

$$h: \mathbf{S}^1 \times [0,1] \to \mathbf{S}^1$$

defined by

$$h(x_1, x_2, t) = (\cos \pi t)(-x_1, -x_2) + (\sin \pi t)(x_2, -x_1)$$

is one such homotopy. In the case of an arbitrary odd n, we have an even number of coordinates and we can again define a homotopy between f and the identity by using a map like h for each pair of coordinates $x_{2i-1}$ and $x_{2i}$, $i = 1,2, \ldots (n + 1)/2$.

Using the properties of the degree of a map, we can easily deduce some well-known theorems. See also Section 8.5.

**7.5.1.4. Theorem.** $\mathbf{S}^n$ is not a retract of $\mathbf{D}^{n+1}$.

**Proof.** Suppose that we have a retraction $r: \mathbf{D}^{n+1} \to \mathbf{S}^n$. Let $\iota: \mathbf{S}^n \to \mathbf{D}^{n+1}$ be the natural inclusion map. The maps

$$\mathbf{S}^n \xrightarrow{\iota} \mathbf{D}^{n+1} \xrightarrow{r} \mathbf{S}^n$$

lead to maps

$$H_n(\mathbf{S}^n) \xrightarrow{\iota_*} H_n(\mathbf{D}^{n+1}) \xrightarrow{r_*} H_n(\mathbf{S}^n).$$

This would imply that the degree of the identity map $r \circ \iota : \mathbf{S}^n \to \mathbf{S}^n$ is zero because $H_n(\mathbf{D}^{n+1}) = 0$. Since the identity map has degree 1 (Theorem 7.5.1.1(1)), the retraction r cannot exist.

More generally, if $\mathbf{W}$ is an orientable pseudomanifold whose boundary $\partial\mathbf{W}$ is nonempty and connected, then $\partial\mathbf{W}$ is not a retract of $\mathbf{W}$. One can use the same argument, but one would have to prove that $H_n(\mathbf{W}) = 0$ first, which involves facts about homology groups that we have not proved. Intuitively, the fact is clear however because there is no nonzero n-cycle since the boundary of the sum of all the n-simplices in a triangulation is nonzero.

**7.5.1.5. Theorem.** (The Brouwer Fixed Point Theorem) Every continuous map $f : \mathbf{D}^n \to \mathbf{D}^n$ has a fixed point.

**Proof.** If f has no fixed points, then the map

$$r : \mathbf{D}^n \to \mathbf{S}^{n-1}$$

defined by

$$r(\mathbf{p}) = \text{point } \mathbf{q} \text{ of } \mathbf{S}^{n-1} \text{ where the ray from } f(\mathbf{p}) \text{ through } \mathbf{p} \text{ meets } \mathbf{S}^{n-1}$$

is a retraction of $\mathbf{D}^n$ onto $\mathbf{S}^{n-1}$, which is impossible by Theorem 7.5.1.4.

Finally,

**Definition.** Let $\mathbf{M}^n$ be a closed orientable pseudomanifold. A homeomorphism $h : \mathbf{M}^n \to \mathbf{M}^n$ is said to be *orientation preserving* if it has degree +1 and *orientation reversing* if it has degree −1.

Note that by Theorem 7.5.1.1(5), the degree of h is ±1. More generally,

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{W}^n$ be closed **oriented** pseudomanifolds and let $f : \mathbf{M}^n \to \mathbf{W}^n$ be a continuous map. As indicated earlier, we can interpret the orientations as corresponding to a choice of generators $\mu_{\mathbf{M}} \in H_n(\mathbf{M}^n)$ and $\mu_{\mathbf{W}} \in H_n(\mathbf{W}^n)$. Define the *degree of f*, denoted by deg f, to be the unique integer defined by the property that

$$f_*(\mu_{\mathbf{M}}) = (\deg f)\mu_{\mathbf{W}}.$$

If f is a homeomorphism, then f is said to be *orientation preserving* if it has degree +1 and *orientation reversing* if it has degree −1.

The degree of a map between pseudomanifolds satisfies properties similar to those stated in Theorem 7.5.1.1. We leave their statements and proofs as exercises for the reader.

### 7.5.2 Manifolds and Poincaré Duality

This section discusses a very important property satisfied by the homology groups of manifolds. It helps greatly in their determination and is a cornerstone in their classification.

Throughout this section we assume that $\mathbf{M}^n$ is a **closed**, **compact**, and **connected** n-dimensional homology manifold and that K is a simplicial complex that triangulates it. Identify the homology groups of **M** with those of K.

First, consider the case where **M** is an **oriented** manifold. The orientation induces a well-defined orientation on all the n-simplices of K. Consider an oriented k-simplex $[\boldsymbol{\sigma}^k]$ and an oriented version of its dual cell $[\boldsymbol{\sigma}_*^{n-k}]$. Both correspond to a union of oriented k-, respectively, (n – k)-simplices of sd(K). Let $\boldsymbol{\nu}$ be any k-simplex of sd(K) contained in $\boldsymbol{\sigma}$ (= $\boldsymbol{\sigma}^k$) and assume that $[\boldsymbol{\nu}]$ has orientation compatible with $[\boldsymbol{\sigma}]$. Similarly, let $\boldsymbol{\tau}_*$ be any (n – k)-simplex of sd(K) in $\boldsymbol{\sigma}_*^{n-k}$ and assume that $[\boldsymbol{\tau}_*]$ has orientation compatible with $[\boldsymbol{\sigma}_*^{n-k}]$. By Proposition 7.5.6(3) $\boldsymbol{\nu}$ and $\boldsymbol{\tau}_*$ have a single vertex in common. Let

$$[\boldsymbol{\nu}] = a[\mathbf{p}_0\mathbf{p}_1 \ldots \mathbf{p}_k] \quad \text{and} \quad [\boldsymbol{\tau}_*] = b[\mathbf{p}_k\mathbf{p}_{k+1} \ldots \mathbf{p}_n],$$

where the $\mathbf{p}_i$ are vertices of sd(K), $\mathbf{p}_k$ is the vertex that $\boldsymbol{\nu}$ and $\boldsymbol{\tau}_*$ have in common, and the integers a and b are ±1. Let

$$[\boldsymbol{\xi}] = c[\mathbf{p}_0\mathbf{p}_1 \ldots \mathbf{p}_n],$$

where the integer c is ±1 and is chosen so that the oriented n-simplex $[\boldsymbol{\xi}]$ has the orientation induced by the given orientation of **M**. The integers a, b, and c clearly depend on how the points $\mathbf{p}_i$ are ordered.

**7.5.2.1. Example.** See Figure 7.33. We have two 2-cells with a counterclockwise orientation. Here n = 2, $\boldsymbol{\sigma} = \mathbf{p}_0\mathbf{p}_3$, $\boldsymbol{\nu} = \mathbf{p}_0\mathbf{p}_1$, k = 1, $\boldsymbol{\sigma}_* = \mathbf{p}_2\mathbf{p}_1 \cup \mathbf{p}_1\mathbf{p}_4$, $\boldsymbol{\tau}_* = \mathbf{p}_2\mathbf{p}_1$, $\boldsymbol{\xi} = \mathbf{p}_0\mathbf{p}_1\mathbf{p}_2$, a = 1, b = c = –1.

**Definition.** The *intersection number* of $[\boldsymbol{\sigma}^k]$ and its dual $[\boldsymbol{\sigma}_*^{n-k}]$, denoted by $I([\boldsymbol{\sigma}^k], [\boldsymbol{\sigma}_*^{n-k}])$, is defined by

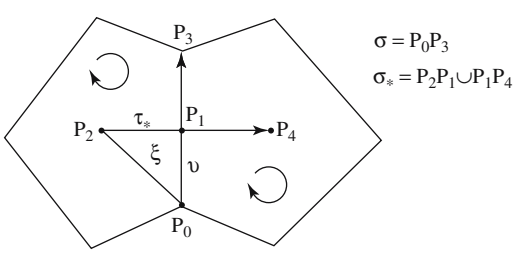



**Figure 7.33.** Defining the intersection number of dual cells.

$$I([\sigma^k],[\sigma_*^{n-k}]) = a\ b\ c.$$

**7.5.2.2. Lemma.** The intersection number $I = I([\sigma^k],[\sigma_*^{n-k}])$ is well defined.

**Proof**. We have to show

(1) I does not depend on the order of the vertices $\mathbf{p}_i$ and
(2) I does not depend on the choice of $\mathbf{v}$ and $\boldsymbol{\tau}_*$.

A proof of these two facts can be found in [SeiT80].

As we pointed out earlier, although the dual cells in the dual complex $K_*$ may not be disks, the fact that we have a homology manifold means that, algebraically, they are linear combinations of simplices that act like disks from the point of view of homology. If one worked through the details one would be able to show that representatives of homology classes can be replaced by linear combinations of the dual cells and that a homology theory that uses chain groups based on dual cells would produce the same homology groups as before. Furthermore, there is also no problem with defining incidence matrices for the dual cells and these would play the same role as the incidence matrices for the simplicial theory.

**7.5.2.3. Theorem.** Given K and $K_*$, if one orients the cells of K and $K_*$ in such a way that the intersection number of dual cells is +1, then the incidence matrix $E^{k-1}$ for K is the transpose of the incidence matrix $E_*^{n-k}$ for $K_*$ multiplied by $(-1)^k$.

**Proof.** See [SeiT80].

**7.5.2.4. Theorem.** (The Poincaré Duality Theorem) The kth Betti number of an orientable homology manifold $\mathbf{M}^n$ is the same as the (n – k)-th Betti number and the k-dimensional torsion coefficients are the same as the (n – k – 1)-dimensional torsion coefficients for k = 0, 1, . . . , n.

**Proof.** See [SeiT80].

Theorem 7.5.2.4 applies only to orientable manifolds. If the manifold is not orientable, then the theorem is false, but one can prove a duality theorem provided one uses $\mathbf{Z}_2$ for coefficients. In that case things get simpler because one does not have to worry about orientations.

**7.5.2.5. Theorem.** (The Mod 2 Poincaré Duality Theorem) The kth connectivity number of a homology manifold $\mathbf{M}^n$ (orientable or not) is the same as the (n – k)-th connectivity number for k = 0, 1, . . . , n.

**Proof.** See [SeiT80].

**7.5.2.6. Corollary.** The Euler characteristic of an odd-dimensional homology manifold is 0.

**Proof.** One simply uses Poincaré duality to show that the terms in the alternating sum cancel each other.

Here is an application of what we know so far.

**7.5.2.7. Theorem.** If $\mathbf{M}^n$ is a simply connected n-dimensional homology manifold with the property that

$$H_i(\mathbf{M}^n) = 0 \quad \text{for } 0 < i \le n/2,$$

then $\mathbf{M}^n$ has the homotopy type of $\mathbf{S}^n$.

**Proof.** First of all, the hypotheses imply that $n \ge 2$. Next, Poincaré duality implies that $H_i(\mathbf{M}) = 0$ for $0 < i < n$. Theorem 7.5.3 and the Hurewicz isomorphism theorem (Theorem 7.4.3.4) then implies that there is a map $f: \mathbf{S}^n \to \mathbf{M}$, so that $[f] \in \pi_n(\mathbf{M})$ gets sent to $\pm 1$ in $H_n(\mathbf{M})$ (which we identified with $\mathbf{Z}$). This map f induces isomorphisms on all homology groups. Since $\mathbf{M}$ is simply connected, f will also induce isomorphisms on the homotopy groups (Theorem 7.4.3.7). Therefore, f is a homotopy equivalence (Theorem 7.4.3.6).

Comparing Theorem 7.5.2.7 with Theorem 7.4.3.8, points out the importance of Poincaré duality. It basically means that we only need to check things on the homology level up to dimension n/2 rather than up to dimension n.

The above discussion of duality in manifolds was really a combinatorial approach to the subject. There are more general approaches. For example, given an n-dimensional homology manifold $\mathbf{M}$, one can define a pairing

$$H_k(\mathbf{M}) \times H_{n-k}(\mathbf{M}) \to \mathbf{Z} \tag{7.11}$$

$$([c],(d)) \to c \bullet d$$

with $c \bullet d$ defined as follows: Choose the representatives c and d for the homology classes so that they intersect transversally, that is, they intersect in a finite number of points. The orientations of the cells of c and d induce an orientation number of $\pm 1$ at these points. Add up these $\pm 1$'s and define $c \bullet d$ to be this sum.

Generalizing further, one can define a pairing

$$H_r(\mathbf{M}) \times H_s(\mathbf{M}) \to H_{r+s-n}(\mathbf{M}) \tag{7.12}$$

$$([c],(d)) \to [c\text{“}\cap\text{”}d]$$

where the representatives c and d are again chosen to intersect transversally and c”$\cap$“d is the cycle defined by the intersection of the simplices of c and d.

In the end though, these geometric approaches to Poincaré duality are “brute force” approaches. The cleanest and most elegant way is via cohomology groups. Here is a brief outline of what one needs to do. First, we need to define something called the cap product. Let $\mathbf{X}$ be a topological space and define a map

$$\cap: \ C^i(\mathbf{X}) \times C_n(\mathbf{X}) \to C_{n-i}(\mathbf{X}) \tag{7.13}$$

as follows: Let $g \in C^i(\mathbf{X})$, $a \in C_n(\mathbf{X})$. Then, using the product of cochains, there is a unique element in $C_{n-i}(\mathbf{X})$, denoted by $g \cap a$, with the property that

$$f(g \cap a) = (f \cdot g)(a), \tag{7.14}$$

for all $f \in C^{n-i}(\mathbf{X})$. In the case of simplicial complexes and an oriented n-simplex $[\boldsymbol{\sigma}]$,

$$g \cap [\boldsymbol{\sigma}] = (-1)^{i(n-i)} g([\text{back } i\text{-face of } \boldsymbol{\sigma}])[\text{front } (n-i)\text{-face of } \boldsymbol{\sigma}], \tag{7.15}$$

where the *front k-face* of $\boldsymbol{\sigma}$ consists of those points with barycentric coordinates $(t_0, t_1, \ldots, t_k, 0, \ldots, 0)$ and the *back k-face* of $\boldsymbol{\sigma}$ consists of those points with barycentric coordinates $(0, \ldots 0, t_{n-k}, \ldots, t_n)$.

**7.5.2.8. Proposition.** The map $\cap$ defined by equation (7.13) is well defined, bilinear, and satisfies

(1) $(g \cdot h) \cap a = g \cap (h \cap a)$
(2) $1 \cap a = a$
(3) $\partial (g \cap a) = (\delta g) \cap a + (-1)^{\dim g}\, g \cap \partial a$

**Proof.** See [MilS74].

Property (3) in Proposition 7.5.2.8 implies that $\cap$ induces a well-defined bilinear map

$$\cap : H^i(\mathbf{X}) \times H_n(\mathbf{X}) \to H_{n-i}(\mathbf{X}) \tag{7.16}$$

**Definition.** The map $\cap$ defined in equation (7.16) is called the *cap product* for $\mathbf{X}$.

**7.5.2.9. Theorem.** (The Poincaré Duality Theorem) Let $\mathbf{M}^n$ be an orientable homology manifold and let $\mu$ be a generator of $H_n(\mathbf{M}) \approx \mathbf{Z}$. The homomorphism

$$\begin{aligned} H^i(\mathbf{M}) &\to H_{n-i}(\mathbf{M}) \\ a &\to a \cap \mu \end{aligned}$$

is an isomorphism for all i.

**Proof.** See [MilS74].

Theorem 7.5.2.9 and Theorem 7.3.1 imply Theorem 7.5.2.4.

## 7.6 Where to Next: What We Left Out

We have covered a lot of algebraic topology, but there is much more and we have only scratched the surface. Of course, we have left out many details and proofs and these should be filled in and understood before moving on, but we would like to mention

some additional topics in this section that would be the natural next step for the interested reader. Since the topics get progressively more advanced, any references for them will necessarily make for harder and harder reading for someone who has only learned about algebraic topology from reading this book. A good general reference is [Span66].

There are a great many tools for computing homology groups. One of the most important is the definition of relative homology groups. If L is a subcomplex of a simplicial complex K, then one can define *relative homology groups* $H_q(K,L)$. These groups are gotten by looking at the groups $C_q(K)/C_q(L)$ and the induced boundary maps

$$\partial_q^r : \frac{C_q(K)}{C_q(L)} \to \frac{C_{q-1}(K)}{C_{q-1}(L)}.$$

Then

$$H_q(K,L) = \frac{\ker \partial_q^r}{\operatorname{im} \partial_{q+1}^r}.$$

(By making the natural definition $C_q(\phi) = 0$, one identifies $H_q(K)$ with $H_q(K,\phi)$.) One can show that, for $q > 0$, $H_q(K,L)$ is isomorphic to $H_q(M)$, where M is a simplicial complex that triangulates the quotient space $|K|/|L|$.

**Definition.** A sequence of abelian groups and homomorphisms

$$\ldots \xrightarrow{h_{q+2}} G_{q+1} \xrightarrow{h_{q+1}} G_q \xrightarrow{h_q} G_{q-1} \xrightarrow{h_{q-1}} \ldots$$

is said to be an *exact sequence* if $\ker h_q = \operatorname{im} h_{q+1}$ for all q.

There is an exact sequence

$$\ldots \to H_q(L) \to H_q(K) \to H_q(K,L) \to H_{q-1}(L) \to H_{q-1}(K) \to \ldots$$

called the *homology sequence of the pair* (K,L) that relates the three homology groups $H_q(L)$, $H_q(K)$, and $H_q(K,L)$, so that if one knows two of the groups, then the third is fairly well determined. This is extremely useful in determining the homology groups of a space from knowledge of the homology groups of subspaces. Simplicial maps on pairs of complexes induce maps on relative homology groups.

One of the problems with simplicial homology theory is that, although one can eventually show that it is a topological invariant, this is not obvious at the start since the groups for a space seem to depend on a particular simplicial subdivision. It would be nicer if one could define groups that are intrinsically topological invariants. Singular homology theory provides the answer. In this theory maps of simplices replace the simplices themselves.

**Definition.** The simplex $\Delta^n = \mathbf{e}_0\mathbf{e}_1 \ldots \mathbf{e}_n$ is called the *standard n-simplex*.

**Definition.** Let **X** be a topological space. A continuous map $T:\Delta^q \to \mathbf{X}$ is called a *singular q-simplex* of **X**. Let $S_q$ be the set of singular simplices of **X**. Define the *group*

*of singular q-chains* of $\mathbf{X}$, denoted by $C_q^s(\mathbf{X})$, by

$$C_q^s(\mathbf{X}) = \{f : S_q \to \mathbf{Z} \mid f(T) = 0 \text{ for all but a finite number of } T \in S_q\}.$$

The group operation "+" on $C_q^s(\mathbf{X})$ is the obvious one, namely, if $f, g \in C_q^s(\mathbf{X})$ and $T \in S_q$, then

$$(f+g)(T) = f(T) + g(T).$$

By identifying T with the map $f_T : S_q \to \mathbf{Z}$, where

$$\begin{aligned} f_T(T) &= 1, \\ f_T(T') &= 0, \text{ if } T' \neq T, \end{aligned}$$

we see (just like in the case of the chain groups for a simplicial complex) that $C_q^s(\mathbf{X})$ can be thought of as the set of all finite linear combinations $n_1T_1 + n_2T_2 + \ldots + n_kT_k$, $n_i \in \mathbf{Z}$, of singular simplices $T_i$ of $\mathbf{X}$.

**Definition.** Given a singular q-simplex $T : \Delta^q \to \mathbf{X}$, define the *ith face* of T,

$$\delta^i T : \Delta^{q-1} \to \mathbf{X}$$

by

$$(\delta^i T)(t_0, t_1, \ldots, t_{q-1}) = T(t_0, \ldots, t_{i-1}, 0, t_i, \ldots, t_{q-1}).$$

The *boundary map*

$$\partial_q^s : C_q^s(\mathbf{X}) \to C_{q-1}^s(\mathbf{X})$$

is the homomorphism defined by the condition that

$$\partial_q^s(T) = \sum_{i=0}^{q} (-1)^i (\delta^i T)$$

for each singular q-simplex $T : \Delta^q \to \mathbf{X}$.

One shows like before that $\partial_{q-1}^s \circ \partial_q^s = 0$, so that one can again define homology groups.

**Definition.** The *qth singular homology group*, denoted by $H_q^s(\mathbf{X})$, is defined by

$$H_q^s(\mathbf{X}) = \frac{\ker \partial_q^s}{\operatorname{im} \partial_{q+1}^s}.$$

**Definition.** Given a continuous map $f : X \to Y$, define a homomorphism

$$f^s_{\#q} : C^s_q(\mathbf{X}) \to C^s_q(\mathbf{Y})$$

by the condition that $f^s_{\#q}(T) = f \circ T$ for every singular q-simplex $T : \Delta^q \to \mathbf{X}$.

One can show that $\partial^s_q \circ f^s_{\#q} = f^s_{\#q-1} \circ \partial^s_q$, so that the maps $f^s_{\#q}$ induce well-defined homomorphisms

$$f^s_{*q} : H^s_q(\mathbf{X}) \to H^s_q(\mathbf{Y}).$$

With the groups $H^s_q(\mathbf{X})$, their corresponding relative groups $H^s_q(\mathbf{X},\mathbf{A})$ for a subspace $\mathbf{A}$ of $\mathbf{X}$, and maps $f^s_{*q}$, along with their relative analogs, we never have to worry about triangulations. The topological invariance is trivially built into the definition. Furthermore, polyhedra now have real groups associated to them, not groups up to isomorphism. The nontrivial part is showing that they give the same groups as the simplicial homology theory. The solution to this problem comes from the fact, referred to earlier in Section 7.2.3, that homology theories can be axiomatized using the Eilenberg-Steenrod axioms. We can take our various definitions of homology groups simply to be existence results that assert that there are objects that satisfy the abstract theory. Any two theories that satisfy the axioms will have isomorphic groups if they have isomorphic homology groups for a point.

Since cohomology groups are derived algebraically from chain groups, one can obviously define *singular cohomology groups*.

Although homotopy groups are much harder to compute than homology groups, there are tools that help in this. One such is the fact that one can define relative homotopy groups $\pi_n(\mathbf{X},\mathbf{A},\mathbf{x}_0)$ that play the same role for homotopy theory that the relative homology groups play for homology theory. Given a topological space $\mathbf{X}$, a subspace $\mathbf{A}$, and a point $\mathbf{x}_0 \in \mathbf{A}$, these groups are obtained from relative homotopy classes of maps

$$\alpha : \left(\mathbf{I}^n, \partial\mathbf{I}^n, \mathbf{e}_1\right) \to (\mathbf{X},\mathbf{A},\mathbf{x}_0),$$

where the homotopies have to keep mapping $\partial\mathbf{I}^n$ to $\mathbf{A}$. Maps between pairs of spaces induce homomorphisms of the relative groups. There is also an exact sequence

$$\ldots \to \pi_{n+1}(\mathbf{X},\mathbf{A},\mathbf{x}_0) \to \pi_n(\mathbf{A},\mathbf{x}_0) \to \pi_n(\mathbf{X},\mathbf{x}_0) \to \pi_n(\mathbf{X},\mathbf{A},\mathbf{x}_0) \to \ldots.$$

From an abstract point of view, we can think of $H_*$ and $\pi_*$ as examples of "functors" from the "category" of topological spaces to the "category" of groups. (The reason for the quotes around some terms is that they have precise mathematical definitions that we cannot go into here.) This is how topological questions get translated into algebraic questions.

The theories we have talked about, homology, homotopy, and so on, really apply to arbitrary topological spaces. Of course manifolds are the most interesting ones, in particular three-dimensional manifolds, because those are the spaces with which we have contact in everyday life. Therefore, it should not be surprising that a great deal of work has been done in low-dimensional topology. Unfortunately, we shall see in the next chapter that, as counter-intuitive as it might seem, a lot more is known about n-dimensional manifolds for $n \geq 5$ than three- and four-dimensional manifolds. As a starting point for more information on advanced aspects of this subject we suggest the books [Mois77] and [Matv03].

Finally, we have said that one major goal of topologists is to find algebraic invariants for spaces that can be used to classify them up to homeomorphism. This chapter has described homology groups, homotopy groups, and cohomology rings that are general-purpose invariants that apply to arbitrary topological spaces. In the next chapter we shall learn about additional invariants that can be defined in the special case of differentiable manifolds. A great many tools for computing the various invariants of this chapter have been developed. We have indicated some of them above. Another extremely powerful tool is the construction of what are called *spectral sequences*, but this subject is much too advanced and technical to even sketch here. The reader would have to have a fairly good understanding of most of the topics described in this chapter beforehand. One good reference is [McCl85].

With regard to the computability of the task of evaluating algebraic invariants, there is unfortunately one negative result along these lines.

**7.6.1. Theorem.** There cannot exist any algorithm for deciding whether or not two given compact, orientable, triangulable 4-manifolds are homeomorphic.

**Proof.** See [Mark58]. The proof depends on the fact that almost any group can be the fundamental group of a 4-manifold and the algebraic fact that the question whether two arbitrary (nonabelian) group representations determine isomorphic groups is undecidable.

Theorem 7.6.1 means that the homeomorphism problem for n-manifolds has a hope of being solvable only when $n \le 3$. The only open case is therefore $n = 3$.

## 7.7 The CW Complex $P^n$

In this last section of the chapter, we return once again to the spaces $\mathbf{P}^n$ and describe them from the point of view of algebraic topology. We begin by defining their standard cell decomposition. This is done best by first describing $\mathbf{S}^n$ as a regular CW complex.

**Definition.** The *standard regular CW complex representation of* $\boldsymbol{S}^n$ is the CW complex defined by the collection of i-dimensional cells $\mathbf{c}_1^i = \mathbf{S}_+^i$ and $\mathbf{c}_2^i = \mathbf{S}_-^i$ in $\mathbf{R}^{i+1}$ for $0 \le i \le n$. The attaching map $f_{i,j}$ for the cell $\mathbf{c}_j^i$ is the obvious one which projects the disks $\mathbf{D}^i$ to either the upper or lower hemisphere of $\mathbf{S}^i$.

Figures 7.8(a) and (c) show the decompositions for $\mathbf{S}^0$ and $\mathbf{S}^1$, respectively. Now think of $\mathbf{P}^n$ as the quotient space of $\mathbf{S}^n$ where we identify antipodal points and let

$$p: \mathbf{S}^n \to \mathbf{P}^n \tag{7.17}$$

be the standard double covering projection.

**Definition.** The *standard regular CW complex representation of* $\boldsymbol{P}^n$ is the CW complex defined by the collection of i-dimensional cells $\mathbf{c}^i = p(\mathbf{c}_1^i)$ and attaching maps $f_i = p \circ f_{i,1}$, for $0 \le i \le n$.

It is worthwhile pointing out the following properties of this CW complex representation of $\mathbf{P}^n$ explicitly:

(1) There is one i-dimensional cell for each dimension $0 \le i \le n$.
(2) The i-skeleton of $\mathbf{P}^n$ is just $\mathbf{P}^i$ and we have a filtration

$$\mathbf{P}^0 = \mathbf{c}^0 \subset \mathbf{P}^1(\approx \mathbf{S}^1) \subset \ldots \subset \mathbf{P}^{n-1} \subset \mathbf{P}^n,$$

where $\mathbf{P}^i$ is obtained from $\mathbf{P}^{i-1}$ by attaching an i-cell.

This cell structure of $\mathbf{P}^n$ allows us to compute the homology and cohomology of the space fairly easily if we use the approach based on oriented cells described in Section 7.2.4. The fact is that each i-cell $\mathbf{c}_1^i$ in $\mathbf{S}^i$ has a natural orientation obtained by projecting the standard orientation of $\mathbf{D}^i$ upward. The projection p projects this orientation to an orientation of the cell $\mathbf{c}^i$ in $\mathbf{P}^n$. If we denote our CW complex for $\mathbf{P}^n$ by C, then since there is only one cell in each dimension i, $0 \le i \le n$,

$$C_i(C) \approx \mathbf{Z}$$

and to compute the homology groups we simply have to analyze the boundary maps on the cells $\mathbf{c}^i$.

**7.7.1. Theorem.** The homology groups of $\mathbf{P}^n$ are given by

(1) $H_0(\mathbf{P}^n) \approx \mathbf{Z}$

$$H_i(\mathbf{P}^n) \approx \begin{cases} 0 & \text{for } 0 < i < n \text{ and i even} \\ \mathbf{Z}_2 & \text{for } 0 < i < n \text{ and i odd} \end{cases}$$

$$H_n(\mathbf{P}^n) \approx \begin{cases} 0 & \text{n even} \\ \mathbf{Z} & \text{n odd} \end{cases}$$

(2) $H_i(\mathbf{P}^n, \mathbf{Z}_2) \approx \mathbf{Z}_2$, $0 \le i \le n$,

**Proof.** Here is a sketch of the argument that proves (1). Let us orient the cells $\mathbf{c}_j^{i+1}$ based on the orientation of $\mathbf{D}^{i+1}$ induced from the standard orientation $[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{i+1}]$ of $\mathbf{R}^{i+1}$. Consider the cell $\mathbf{c}_1^{i+1}$ in $\mathbf{S}^n$ with $i \ge 2$. The boundary of that cell consists of the two cells $\mathbf{c}_1^i$ and $\mathbf{c}_2^i$. What orientation does the boundary map on $\mathbf{c}_1^{i+1}$ induce on these two cells? Well, the orientation induced on $\mathbf{c}_1^i$ and $\mathbf{c}_2^i$ have to be $[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_i]$ and $[-\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_i]$, respectively. The reason is that adding $\mathbf{e}_{i+1}$ at the end of the first basis and $-\mathbf{e}_{i+1}$ at the end of the second must lead to the standard orientation of $\mathbf{D}^{i+1}$ and it is easy to check in the second case that

$$[-\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_i, -\mathbf{e}_{i+1}] = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{i+1}].$$

But under the antipodal identification map the cell $\mathbf{c}_2^i$ and its orientation $[-\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_i]$ gets mapped to the cell $\mathbf{c}_1^i$ with orientation $[\mathbf{e}_1, -\mathbf{e}_2, \ldots, -\mathbf{e}_i]$. The latter agrees

with the orientation $[\mathbf{e}_1,\mathbf{e}_2, \ldots, \mathbf{e}_i]$ if i is odd and $-[\mathbf{e}_1,\mathbf{e}_2, \ldots, \mathbf{e}_i]$ if i is even. In the first case, $\partial\mathbf{c}_j^{i+1}$ will equal $2\mathbf{c}_1^i$ and in the second, 0. What this shows is that, for $0 < i \leq n$,

$$\begin{aligned} Z_i(C) &= 0 \quad \text{if i is even} \\ &\approx \mathbf{Z} \quad \text{if i is odd} \end{aligned}$$

and, for $0 < i < n$,

$$\begin{aligned} B_i(C) &= 0 \quad \text{if i is even} \\ &\approx 2\mathbf{Z} \quad \text{if i is odd.} \end{aligned}$$

Part (2) of the theorem follows from the fact that 2 is the same as 0 in $\mathbf{Z}_2$. For a more rigorous proof of this theorem see [CooF67].

**7.7.2. Theorem.** The cohomology groups and cohomology ring structure of $\mathbf{P}^n$ are:

(1) $H^0(\mathbf{P}^n) \approx \mathbf{Z}$

$$H^i(\mathbf{P}^n) \approx \begin{cases} 0 & \text{for } 0 < i < n \text{ and i odd} \\ \mathbf{Z}_2 & \text{for } 0 < i < n \text{ and i even} \end{cases}$$

$$H^n(\mathbf{P}^n) \approx \begin{cases} \mathbf{Z}_2 & \text{n even} \\ \mathbf{Z} & \text{n odd} \end{cases}$$

(2) $H^i(\mathbf{P}^n,\mathbf{Z}_2) \approx \mathbf{Z}_2$, $0 \leq i \leq n$,
(3) As a ring using the cup product, $H^*(\mathbf{P}^n,\mathbf{Z}_2)$ is a polynomial ring with one generator $w_n \in H^1(\mathbf{P}^n,\mathbf{Z}_2)$ satisfying a single relation $w_n^{n+1} = 0$, that is,

$$H^*(\mathbf{P}^n,\mathbf{Z}_2) \text{“}\approx\text{”} \frac{\mathbf{Z}_2[w_n]}{(w_n^{n+1})}.$$

If $i: \mathbf{P}^n \subset \mathbf{P}^{n+1}$ is the natural inclusion, then $i^*(w_{n+1}) = w_n$.

**Proof.** See [CooF67] and [Span66]. We could use Theorem 7.3.1 for part (1).

We finish with a result about the homotopy groups of $\mathbf{P}^n$. Recall, however, our earlier comment that the sphere $\mathbf{S}^n$ has nontrivial higher homotopy groups.

**7.7.3. Theorem.** Let $n \geq 1$.

(1) Since $\mathbf{P}^1$ is homeomorphic to $\mathbf{S}^1$, $\pi_1(\mathbf{P}^1) \approx \mathbf{Z}$.
(2) If $n > 1$, then $\pi_1(\mathbf{P}^n) \approx \mathbf{Z}_2$.
(3) $\pi_i(\mathbf{P}^n) \approx \pi_i(\mathbf{S}^n)$ for $i \geq 2$.

**Proof.** See [Stee51] or [Span66]. The isomorphisms in (3) are induced by the projection map in (7.17).

## 7.8 EXERCISES

### Section 7.2.1

7.2.1.1. Prove Lemma 7.2.1.2.

7.2.1.2. Prove that if K is a simplicial complex, then

$$\text{rank } (H_0(K)) = \text{number of connected components of } |K|.$$

7.2.1.3. Prove that the homology groups of the Klein bottle are as indicated in Table 7.2.1.1. You can use a triangulation similar to that of the torus shown in Figure 7.5.

7.2.1.4. Prove the results indicated in Table 7.2.1.1 for

(a) orientable surfaces of genus k
(b) nonorientable surfaces of genus k

7.2.1.5. If **X** and **Y** are polyhedra, prove that

$$H_q(\mathbf{X} \vee \mathbf{Y}) \approx H_q(\mathbf{X}) \oplus H_q(\mathbf{Y}), \quad q \neq 0.$$

### Section 7.2.2

7.2.2.1. Prove Lemma 7.2.2.6.

7.2.2.2. Let $K = \partial\langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\rangle$ be the simplicial complex in Example 7.2.1.6. Compute the maps $f_{*q}$ for the simplicial map $f: K \to K$ defined by $f(\mathbf{v}_0) = \mathbf{v}_0$, $f(\mathbf{v}_1) = \mathbf{v}_2$, $f(\mathbf{v}_2) = \mathbf{v}_1$, and $f(\mathbf{v}_3) = \mathbf{v}_3$.

7.2.2.3. Let K be a simplicial complex. Show that sd(K) is also a simplicial complex.

### Section 7.2.3

7.2.3.1. Let K be a simplicial complex. Show that the cone on $|K|$ is a polyhedron and determine its homology groups.

7.2.3.2. If K is a nonempty simplicial complex and if $|K|$ has k components, prove that the suspension of $|K|$ has the following homology groups:

$$\begin{aligned} &H_0(S|K|) = \mathbf{Z} \\ &H_1(S|K|) \approx (k-1)\mathbf{Z} \\ &H_{q+1}(S|K|) \approx H_q(K), \quad q > 0. \end{aligned}$$

**Section 7.2.4**

7.2.4.1. Describe a minimal cell decomposition for the Klein bottle.

7.2.4.2. Triangulate the dunce hat and compute its homology groups.

7.2.4.3. (a) Prove that the homology groups of the lens spaces are what they were stated to be.
(b) Prove that the Euler characteristic of a lens space is 0.

**Section 7.2.5**

7.2.5.1. Compute the incidence matrices for the simplicial complex $K = \partial\langle \mathbf{v}_0\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\rangle$. Work through the proof of Theorem 7.2.5.3 and determine the normalized form of the incidence matrices and the bases of the chain groups that define them. Show how these matrices determine the known homology groups.

**Section 7.2.6**

7.2.6.1. (a) If $\mathbf{X}$ is a point, then prove that

$$H_0(\mathbf{X};G) \approx G, \text{ and}$$
$$H_q(\mathbf{X};G) = 0, \text{ for } q \neq 0.$$

(b) Let $\mathbf{X}$ be a polyhedron. Prove that $H_0(\mathbf{X};G)$ is isomorphic to a direct sum of as many copies of G as there are components of $\mathbf{X}$. In particular, the 0th connectivity number of $\mathbf{X}$, $\kappa_0(\mathbf{X})$, is nothing but the number of components of $\mathbf{X}$.

7.2.6.2. Compute $H_q(\mathbf{X};\mathbf{Z}_2)$, for all q, where $\mathbf{X}$ is

(a) $\mathbf{S}^n$
(b) $\mathbf{S}^1 \times \mathbf{S}^1$
(c) $\mathbf{P}^2$
(d) the Klein bottle

**Section 7.3**

7.3.1. Prove that the space $\mathbf{X} = \mathbf{S}^2 \vee \mathbf{S}^1 \vee \mathbf{S}^1$ in Figure 7.15 has the same homology groups as the torus.

**Section 7.4.1**

7.4.1.1. Complete the proof of Theorem 7.4.1.6 by filling in all missing details.

7.4.1.2. Prove Theorem 7.4.1.7.

7.4.1.3. Prove Lemma 7.4.1.11.

7.4.1.4. Prove Theorem 7.4.1.12.

**Section 7.4.2**

7.4.2.1. Show that the bundle $(\mathbf{S}^n, p)$ over $\mathbf{P}^n$ in Example 7.4.2.1 is locally trivial.

7.4.2.2. Give an intuitive justification of the fact that the lens spaces defined in Sections 7.2.4 and 7.4.2 are homeomorphic.

**Section 7.4.3**

7.4.3.1. Prove Theorem 7.4.3.1.

7.4.3.2. Prove Lemma 7.4.3.3.

**Section 7.5**

7.5.1. Work through the details of a proof of Theorem 7.5.1.

7.5.2. Work through the details of a proof of Proposition 7.5.6.

**Section 7.5.1**

7.5.1.1. Let $f, g: \mathbf{S}^n \to \mathbf{S}^n$, $n \geq 1$, be continuous maps.

(a) Prove that if $f(\mathbf{p}) \neq g(\mathbf{p})$ for all $\mathbf{p} \in \mathbf{S}^n$, then

$$\deg f + (-1)^n (\deg g) = 0.$$

(b) Prove that if $\deg f \neq (-1)^{n+1}$, then f has a fixed point.

(c) Prove that if $\deg f \neq 1$, then $f(\mathbf{p}) = -\mathbf{p}$ for some $\mathbf{p} \in \mathbf{S}^n$.

CHAPTER 8

# Differential Topology

## 8.1 Introduction

Most of what we did in the last three chapters applied to topological spaces that could be quite general, even if one restricted oneself to polyhedra. In this chapter we specialize to studying manifolds. Topological manifolds were defined in Chapter 5 and then studied further in the context of pseudomanifolds and homology manifolds in Chapter 7. As topological spaces they look like $\mathbf{R}^n$ or $\mathbf{R}^n_+$ locally. To put it another way, to a sufficiently small bug in a manifold the space around it would look flat. Now we shall study differentiable manifolds, which have a differentiable structure in addition to their topology. Having a differentiable structure on a manifold means that we can transfer many other properties and techniques from Enclidean space over to it. In particular, we can use calculus and linear algebra. This turns out to play an important role in the analysis of the manifold because we will be able to do things that we can not do with an ordinary topological manifold. Smooth curves and surfaces, spaces important in geometric modeling and CAGD, are instances of differentiable manifolds.

Although we sketch an intrinsic definition of a differentiable manifold later in this chapter, our working definition will be one that defines certain subspaces of $\mathbf{R}^n$ to be manifolds. By defining a manifold in terms of what could be considered as a particular imbedding of the corresponding "abstract" manifold, we simplify the definition of tangent vector and tangent space, which are an essential part of a manifold. The intrinsic definition of a manifold is not that much harder, but with our approach we make it somewhat easier for the reader who has never seen any of this material before and who does not feel entirely comfortable with n dimensions. Defining manifolds as parameterized subspaces of Euclidean space is also the way one usually sees them defined in CAGD. In one sense, there is no loss of generality because an important theorem asserts that even abstract manifolds can be realized as subspaces of a suitably high dimensional Euclidean space. On the other hand, one should be aware of the fact that the disadvantage of studying manifolds as subsets of Euclidean space is that it might seem as if some of the invariants we define for them depend on the surrounding space when in fact they do not. Abstraction enables one to see essential

aspects more clearly. That is the reason that one defines an abstract n-dimensional vector space and does not just deal with $\mathbf{R}^n$ (although the two are isomorphic). In the case of manifolds, they have lots of intrinsic properties that do not depend on any particular imbedding. For example, on one level anyway, there really is no difference between all the circles of radius one in the plane. They all correspond to different imbeddings of an "abstract" version of a unit circle. In fact, until we get to differential geometry where the metric matters, **all** circles are the "same."

Section 8.2 discusses parameterizations of spaces, which is essential for the definition of a manifold given in Section 8.3 and the abstract manifolds defined in Section 8.8. The idea of a manifold originated in Riemann's groundbreaking lecture "On the Hypotheses which lie at the Foundation of Geometry" delivered to the faculty at the University of Göttingen in 1854. The ideas expressed in this talk are usually considered to be the most influential in the history of differential geometry. An integral part of a differentiable manifold is its tangent space, which is defined in Section 8.4. Section 8.5 discusses what it means for a manifold to be orientable. Sections 8.6 and 8.7 give an overview of what is involved in the classification of manifolds. They give the reader a taste of some difficult but beautiful results on the structure of manifolds. Key to this is the handle decomposition of a manifold and cobordism theory along with algebraic topology invariants. This part of the theory is relatively new. It covers about a twenty year period starting in the middle 1950s and culminated in the main structure theorems for manifolds that are known today. See [AgoM76b]. Next, in Section 8.8 we move on to an intrinsic definition of a manifold. Sections 8.9 and 8.10 define vector bundles and discuss some of their basic properties with emphasis on their role in the study of manifolds. Section 8.11 defines what it means for maps or manifolds to be transverse. The degree of a map and intersection numbers serve as two examples of how transversality can be used, but it appears in a great many essential ways in proofs related to differentiable manifolds. In Section 8.12 we continue the topic of differential forms and integration that we started in Sections 4.9 and 4.9.1, but this time in the setting of manifolds. Finally, in Section 8.13 we take another look at the projective spaces $\mathbf{P}^n$ and we introduce the Grassmann manifolds in Section 8.14.

## 8.2 Parameterizing Spaces

Parameterizations are a generalization of Descartes' idea that a good approach to studying a geometric space is to introduce coordinates for its points, because one can then use equations or other analytic tools to study the space. They are used in many places and are fundamental to the idea of a manifold, especially the abstract manifolds defined at the end of this chapter.

**Definition.** Let $\mathbf{X}$ be a subset of $\mathbf{R}^n$. If there a subset $\mathbf{U}$ of $\mathbf{R}^k$ and a surjective $C^r$ map

$$\Phi\colon \mathbf{U} \to \mathbf{X},$$

then $\Phi$ is called a $C^r$ *parameterization* of $\mathbf{X}$. The set $\mathbf{X}$ is called the *underlying space* of $\Phi$ in that case.

Usually, parameterizations will be one-to-one maps, or close to that. In that case, they basically allow us to associate coordinates to points of a space, that is, if

$$\Phi(u_1, u_2, \ldots, u_k) = \mathbf{q},$$

then we can think of $\mathbf{q}$ as having coordinates $(u_1, u_2, \ldots, u_k)$. (In the context of parametrizations and a set $\mathbf{U}$ we shall often use parameters $u_i$ rather than $x_i$.) The modifier "$C^r$" is often omitted and may need to be determined from the context. As a general rule, unless explicitly stated otherwise, *the assumption will be that a parameterization is a $C^\infty$ map*. The reason is that we frequently want to have the ability to talk about the derivative of the map. Actually, to be of class $C^3$ would be adequate for everything we do in this book but we do not want to get involved in that sort of technical detail. In this context, for differentiability to make sense, $\mathbf{U}$ will be either an open subset of $\mathbf{R}^k$ or have a "nice" boundary, that is, every boundary point will have a neighborhood that looks like $\mathbf{R}^k_+$.

**8.2.1. Example.** The graph of any function admits a natural parameterization. More precisely, if $\mathbf{A} \subseteq \mathbf{R}^m$ and if $f: \mathbf{A} \to \mathbf{R}^n$, then the map

$$\Phi\colon\ \mathbf{A} \to \text{graph}(f) \subseteq \mathbf{R}^m \times \mathbf{R}^n$$

defined by

$$\Phi(\mathbf{a}) = (\mathbf{a}, f(\mathbf{a})) \tag{8.1}$$

is a parameterization of the graph of f. One can think of this map as projecting orthogonally up from $\mathbf{A}$ to the graph. Here are two examples:

$$\begin{aligned} f(x) = x^2: &\qquad \Phi(x) = (x, x^2) \\ f(x,y) = x^2 + y^2: &\quad \Phi(x,y) = (x, y, x^2 + y^2) \end{aligned}$$

See Figure 8.1. In differential geometry the parameterization $\Phi(x,y) = (x,y,f(x,y))$ of the surface that is the graph of the function $f(x,y)$ is often called a *Monge patch*.

To get a parameterization for a space, one usually has to think of some geometric way that its points can be described by some real numbers. These real numbers correspond to "directions" as to "how one can get to the point." For example, the coordinates of the point (2,3) in $\mathbf{R}^2$ can be thought of as saying that one can get to it from the origin by walking a distance 2 along the x-axis and then a distance 3 along a line parallel to the y-axis.

**8.2.2. Example.** Consider the unit sphere $\mathbf{S}^2$ as being a surface of revolution obtained by rotating one half of the unit circle about the x-axis. By definition of a surface of revolution, the sphere is then a union of circles each of which is the intersection of the surface with a plane parallel with the yz-plane that meets the x-axis at some x-value. But any one of these circles is, again by definition, just a point revolved about the x-axis. Since points on a circle of fixed radius can be specified by one real

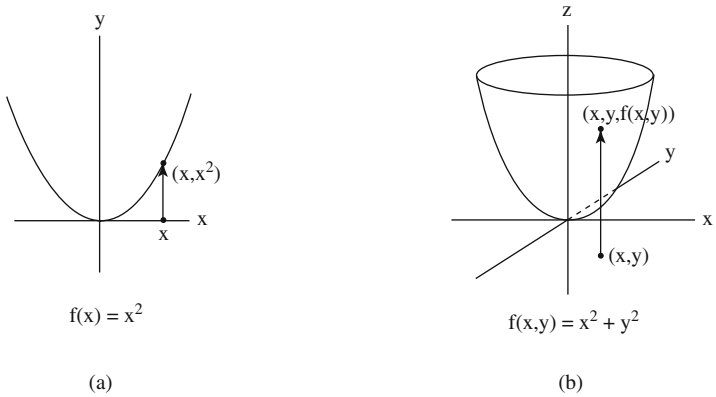



**Figure 8.1.** Parameterizations of graphs.

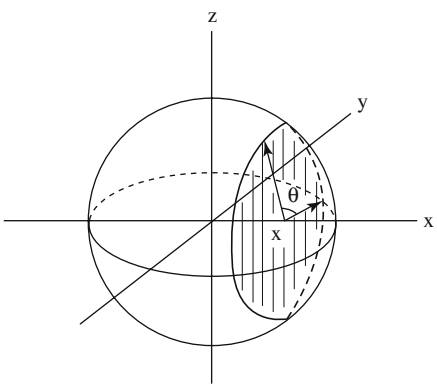



**Figure 8.2.** A parameterization for a sphere.

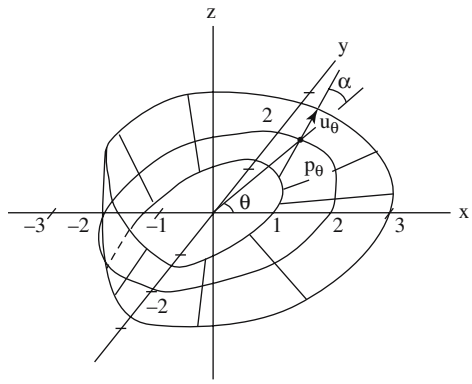
number, an angle, it is therefore clear that one could tell someone how to get to a point on the sphere by telling that person two numbers, x and θ. The x-value specifies a circle and its radius and the θ-value a point on that circle. This leads to the parameterization of $\mathbf{S}^2$ defined by

$$\Phi(x,\theta) = \left(x, \sqrt{1-x^2}\cos\theta, \sqrt{1-x^2}\sin\theta\right), \quad -1 \le x \le 1, 0 \le \theta \le \pi. \tag{8.2}$$

See Figure 8.2. Note that $\sqrt{1-x^2}$ is the radius of the circle at x.

**8.2.3. Example.** A slightly more complicated example is a parameterization of the Moebius strip. See Figure 8.3. One can think of the Moebius strip as a union of line segments parameterized by [–1,1], one for each point on the circle of radius 2 about the origin. Now an ordinary vertical cylinder centered on the z-axis of radius 2 could be parameterized by

**Figure 8.3.** Parameterizing the Moebius strip.




$$\Phi(\theta, t) = (2\cos\theta, 2\sin\theta, t), \quad 0 \le \theta < 2\pi \quad \text{and} \quad -1 \le t \le 1,$$

but in the case of the Moebius strip we need to rotate the line segments about their center point. When θ is zero, we start with a horizontal line segment from 1 to 3 on the x-axis. As θ increases, we start rotating the line segment about its center on the circle of radius 2 with the top end tilting up toward the z-axis. When θ gets to π, the line segment is vertical. The parameterization is easy to write in vector form. Let

$$\mathbf{e}_\theta = (\cos\theta, \sin\theta, 0)$$

and let $\mathbf{u}_\theta$ denote the unit vector which, at "time" θ, points from $\mathbf{p}_\theta = 2\mathbf{e}_\theta$ to the point at the top of the vertically-slanted line segment. Then the parameterization we want is

$$\Phi(\theta, t) = 2\mathbf{e}_\theta + t\mathbf{u}_\theta.$$

It remains to compute $\mathbf{u}_\theta$ in terms of θ and t. But $\mathbf{u}_\theta$ lies on the unit circle in the vertical plane through the origin with basis $\mathbf{e}_\theta$ and $\mathbf{e}_3$. If it makes an angle α with the vector $\mathbf{e}_\theta$, then

$$\mathbf{u}_\theta = \cos\alpha\, \mathbf{e}_\theta + \sin\alpha\, \mathbf{e}_3.$$

Since $\alpha = \theta/2$, we are led to the following formula for Φ:

$$\Phi(\theta, t) = \left(\left(2 + t\cos\frac{\theta}{2}\right)\cos\theta, \left(2 + t\sin\frac{\theta}{2}\right)\sin\theta, t\sin\frac{\theta}{2}\right), \tag{8.3}$$

where $0 \le \theta < 2\pi$ and $-1 \le t \le 1$.

Parameterizations are intended to help in the study of a space. They are usually not of interest by themselves. The idea is that by determining properties of the map one gets some information about intrinsic properties of the space. As mentioned earlier, the definition as it stands is really too general to expect something like this to work unless the map is essentially one-to-one.

**Definition.** Let $\mathbf{U} \subseteq \mathbf{R}^k$. A $C^r$ map, $r \geq 1$,

$$\Phi\colon\ \mathbf{U} \to \mathbf{R}^n$$

is said to be *regular at a point* $\boldsymbol{p}$ *in* $\boldsymbol{U}$ if $D\Phi(\mathbf{p})$ is one-to-one. The map $\Phi$ is said to be *regular* if it is regular at every point of $\mathbf{U}$.

The parameterization in Example 8.2.1 of the graph of a function is regular, assuming that the function is differentiable. The parameterization in Example 8.2.3 is also regular. Both are in fact globally one-to-one. In general, regular parameterizations are locally one-to-one (Theorem 4.4.6), but not necessarily globally one-to-one.

**8.2.4. Example.** The parameterization

$$\Phi(\theta) = (\cos\theta, \sin\theta),\ \theta \in \mathbf{R},$$

of the unit circle is a regular parameterization but is not globally one-to-one.

The parameterization of $\mathbf{S}^2$ defined in Example 8.2.2, although often used, is not regular. For one thing, it is not differentiable when x is ±1. For another, when x is ±1 the circle being rotated has degenerated to a point and $\Phi$ is not locally one-to-one there. The nonregularity may, however, not be a problem if one is not interested in those values of x.

Even if one sticks to regular parameterizations, there are still many ways to parameterize a space. For example, if a curve is parameterized by an interval and one thinks of the parameter as time, then one can traverse or walk along the curve with many different velocities and each one would correspond to a different parameterization of the curve. When using parameterizations as a vehicle for studying spaces we must be careful to stick to those properties that are an invariant of the underlying space.

**Definition.** Let $\mathbf{U}, \mathbf{V} \subseteq \mathbf{R}^k$ and $\mathbf{X} \subseteq \mathbf{R}^n$. Let $r \geq 1$ and let $\Phi\colon \mathbf{U} \to \mathbf{X}$ and $\Psi\colon \mathbf{V} \to \mathbf{X}$ be two regular $C^r$ parameterizations of a space $\mathbf{X}$. We say that $\Psi$ is a *regular reparameterization* of $\Phi$ if $\Psi = \Phi \circ \mu$ for some one-to-one and onto $C^r$ map $\mu\colon \mathbf{V} \to \mathbf{U}$ with $D\mu(\mathbf{q})$ one-to-one for all $\mathbf{q}$ in $\mathbf{V}$. The map $\mu$ will be called a *change of coordinates* or *change of parameters transformation*. The map $\mu$ is said to be *orientation preserving* if $\det(D\mu(\mathbf{q})) > 0$ for all $\mathbf{q}$ in $\mathbf{V}$; otherwise, $\mu$ is said to be *orientation reversing*. If $\mu$ exists, then $\Phi$ and $\Psi$ are said to be *equivalent* parameterizations.

See Figure 8.4. Think of the map $\mu$ as defining a change of coordinates. We shall sometimes say that $\Psi$ was obtained from $\Phi$ by a change in coordinates. The properties of spaces that we want to study via parameterizations should be invariant under regular reparameterizations. It is easy to show that the notion of being equivalent is an equivalence relation on the set of regular parameterizations of a set $\mathbf{X}$.

**8.2.5. Example.** Consider the parameterizations

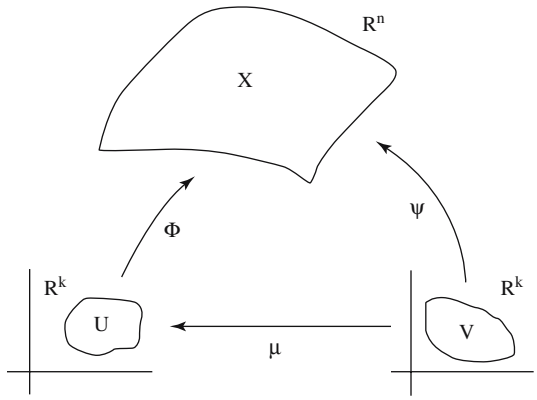



**Figure 8.4.** Regular reparameterizations.

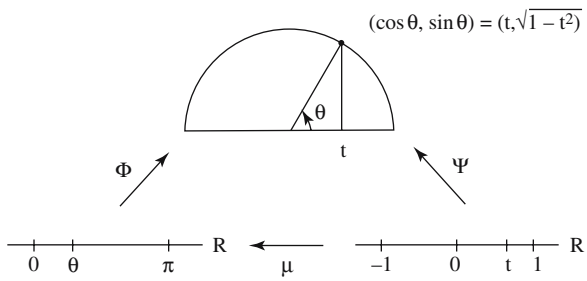



**Figure 8.5.** Reparameterizing the half circle.

$$\mathbf{\Phi}\colon\ [0,\pi] \to \mathbf{S}^1_+ \quad \text{and} \quad \Psi\colon\ [-1,1] \to \mathbf{S}^1_+$$

defined by

$$\Phi(\theta) = (\cos\theta, \sin\theta) \quad \text{and} \quad \Psi(t) = \left(t, \sqrt{1-t^2}\right).$$

The map

$$\mu\colon\ [-1,1] \to [0,\pi],\ \mu(t) = (1-t)\frac{\pi}{2},$$

is an orientation-reversing reparameterization because $\mu'(t) = -\pi/2$. See Figure 8.5.

## 8.3 Manifolds in $R^n$

Topological manifolds were defined in Section 5.3. Now we add a differential structure. In this section we restrict ourselves to subsets of Euclidean space because, by

doing so, the definitions really only involve concepts that should be familiar to the reader, such as the differentiability of vector-valued functions. The price we pay, however, is that they are not entirely satisfactory from a mathematical point of view. For one thing, the reader will find that we never really define a differential structure anywhere in this section. We shall only be defining whatever is needed for two ideas to make sense, namely, that a manifold has tangent planes and that certain functions are differentiable. The correct and intrinsic definitions are postponed to Section 8.8 at which point the reader has hopefully gotten a feeling for the geometric ideas, so that the additional abstraction will not be a problem.

We again phrase our definitions in a way that includes manifolds with boundary right at the start. The reader should compare the new definition with the one in Section 5.3.

**Definition.** A subset $\mathbf{M}$ of $\mathbf{R}^n$ is called a *k-dimensional $C^r$ manifold*, $r \geq 0$, if, for every point $\mathbf{p}$ in $\mathbf{M}$, there is an open neighborhood $\mathbf{V_p}$ of $\mathbf{p}$ in $\mathbf{M}$, an open set $\mathbf{U_p}$ in $\mathbf{R}^k_+$, and a $C^r$ homeomorphism $\Phi_\mathbf{p}: \mathbf{U_p} \to \mathbf{V_p}$, which is assumed to be regular if $r \geq 1$. The maps $\Phi_\mathbf{p}$ are called *local ($C^r$) parameterizations* for $\mathbf{M}$. A $C^\infty$ manifold is called simply a *differentiable or smooth manifold*. If $\mathbf{V_p} = \mathbf{M}$, then $\Phi_\mathbf{p}$ is called a *proper ($C^r$) parameterization* for $\mathbf{M}$. The *boundary* of the manifold $\mathbf{M}$, $\partial\mathbf{M}$, is defined by

$$\partial\mathbf{M} = \left\{\mathbf{p} \in \mathbf{M} \,\middle|\, \Phi_\mathbf{p}^{-1}(\mathbf{p}) \in \mathbf{R}^{k-1}\right\}.$$

If $\partial\mathbf{M} = \mathbf{\Phi}$, then $\mathbf{M}$ is called a *closed* manifold.

See Figure 8.6. Clearly, every $C^r$ manifold is also a $C^s$ manifold for $0 \leq s \leq r$. Since a $C^0$ manifold is almost by definition a topological manifold, it follows that every $C^r$ manifold is a topological manifold and so the terminology used with the latter applies. Furthermore, from what we know about topological manifolds it follows that the dimension and boundary of $C^r$ manifolds are well defined. It is also easy to show that the boundary of a k-dimensional $C^r$ manifold is a $(k - 1)$-dimensional $C^r$ manifold (without boundary). One slight difference between this definition and the earlier one in Section 5.3 is that we have used an arbitrary open set $\mathbf{U_p}$ in $\mathbf{R}^k_+$ rather than just the whole halfspace $\mathbf{R}^k_+$ One does not gain any generality by doing so, but it will match

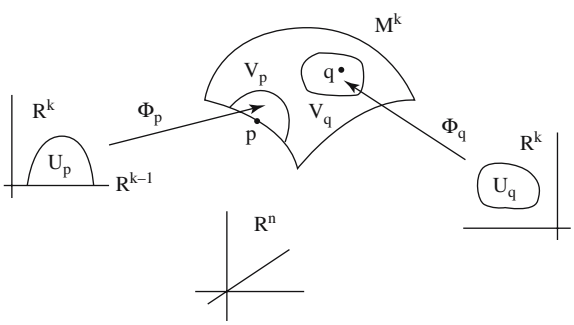



**Figure 8.6.** A k-dimensional manifold $\mathbf{M}^k$.

the definitions in Section 8.8 better. The reason for using sets $\mathbf{U_p}$ will become apparent later in Section 8.8.

**Note.** Trying to handle both closed manifolds and manifolds with boundary does introduce some complications because results involving the latter sometimes need special hypotheses. One difference shows up when it comes to limits or derivatives. In the future, such limits or derivatives at points on the boundary should always be taken to mean one-sided limits or derivatives although we shall not explicitly say so. With that standing assumption, theorems that we state for manifolds will hold for both types unless we say otherwise.

We shall see that one important difference between a subset of $\mathbf{R}^n$ being a topological manifold and being a differentiable manifold is that a subset that is a differentiable manifold has a nice, unique "tangent plane" at every point. Think of a sphere and its tangent planes. A precise definition of tangent planes will be given in the next section. Their existence is a consequence of the regularity of the local parameterizations. For that reason, the regularity property of local parameterizations is an essential hypothesis and not just a minor property that has been tacked on to the definition. This means that, although the boundary of a square is a topological 1-manifold, it is not a $C^r$-manifold, $r \geq 1$, because there is no unique tangent line at the corners. A more correct way of stating this fact is to say that the boundary is not a differentiable **sub**manifold of $\mathbf{R}^n$. In general, this section simply specifies sufficient conditions for a subset to be a differentiable submanifold of $\mathbf{R}^n$, namely, a differentiable imbedding of an abstract manifold as defined in Section 8.8. It should also be pointed out that if a set $\mathbf{M}$ is a $C^r$ manifold, then we have assumed the existence of certain local $C^r$ parameterizations, but it does **not** follow that **every** local parameterization of $\mathbf{M}$ will be a $C^r$ parameterization. See Example 8.3.2 below.

**8.3.1. Example.** Euclidean space $\mathbf{R}^k$ is a k-dimensional differentiable manifold because we can let $\mathbf{U_p} = \mathbf{R}^k$ and let $\Phi$ be the identity map. It is also easy to show that $\mathbf{D}^k$ is k-dimensional differentiable manifold with boundary $\mathbf{S}^{k-1}$.

**8.3.2. Example.** The unit circle $\mathbf{S}^1$ is a one-dimensional differentiable manifold because it can be covered by local $C^\infty$ parameterizations

$$\Phi_i : (-1,1) \to \mathbf{S}^1, \quad i = 1,2,3,4,$$

defined by

$$\begin{aligned}
\Phi_1(u) &= \left(u, \sqrt{1-u^2}\right), \\
\Phi_2(u) &= \left(\sqrt{1-u^2}, u\right), \\
\Phi_3(u) &= \left(u, -\sqrt{1-u^2}\right), \quad \text{and} \\
\Phi_4(u) &= \left(-\sqrt{1-u^2}, u\right).
\end{aligned}$$

Each map $\Phi_i$ covers half of the circle and is a $C^\infty$ map since we have stayed away from $u = \pm 1$. The maps

$$\Psi_1 : \left(0, \frac{3\pi}{2}\right) \to \mathbf{S}^1 \quad \text{and} \quad \Psi_2 : \left(-\pi, \frac{\pi}{2}\right) \to \mathbf{S}^1$$

defined by

$$\Psi_1(u) = \Psi_2(u) = (\cos u, \sin u)$$

are a simpler set of local $C^\infty$ parameterizations. There are lots of different local $C^\infty$ parameterizations for $\mathbf{S}^1$. On the other hand, there are also lots of local parameterizations that are not $C^\infty$. For example the map

$$\varphi : (-1,1) \to \mathbf{S}^1_+$$

defined by

$$\begin{aligned} \varphi(u) &= \left(-\sqrt{1-(1+u)^2}, 1+u\right), u \in (-1,0], \\ &= \left(\sqrt{1-(1-u)^2}, 1-u\right), u \in [0,1), \end{aligned}$$

is parameterization of a neighborhood of the point (0,1) that is continuous ($C^0$) but not $C^1$ because the derivative does not exist at $u = 0$.

The next theorem states an important property of differentiable manifolds, although its proof is much too involved to present here. The main consequence for us is that simplicial homology groups are defined for such manifolds and we can use what we know about pseudomanifolds and homology manifolds.

**8.3.3. Theorem.** Every $C^r$ manifold $\mathbf{M}$, $r \geq 1$, admits a triangulation that is infinite in general, but if $\mathbf{M}$ is compact, then $\mathbf{M}$ has a finite triangulation that makes it into a pseudomanifold if it is connected. Every closed compact connected $C^r$ manifold, $r \geq 1$, is a homology manifold.

**Proof.** See [Munk61].

Manifolds are defined in terms of parameterizations and that is the most common way they are presented, but there is another way, namely, they can sometimes be defined as the set of zeros of a function. For example, the sphere $\mathbf{S}^2$ is the set of zeros of the polynomial

$$p(x,y,z) = x^2 + y^2 + z^2 - 1.$$

**Definition.** Let $f : \mathbf{R}^n \to \mathbf{R}^m$. Define the set of zeros of f, V(f), by

$$V(f) = f^{-1}(\mathbf{0}) = \{\mathbf{p} \in \mathbf{R}^n \mid f(\mathbf{p}) = 0\}.$$

In practice, f is usually a polynomial, and in that case V(f) is also called an (affine algebraic) *variety*. Algebraic geometry is that field in mathematics which tries to analyze the topological structure of V(f) in terms of algebraic invariants associated to


**Figure 8.7.** Surfaces that are varieties.

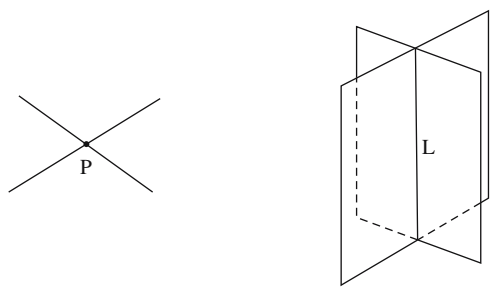



**Figure 8.8.** A variety that is not a surface.

the polynomial f. It happens that there are some very deep connections between the two. We shall look at some of this in Chapter 10. The graphs of functions can always be thought of as varieties. For example, in Figure 8.7 the graph of the functions

$$z = g(x, y)$$

are varieties V(f), where

$$f(x, y, z) = z - g(x, y).$$

Not all zero sets V(f) are nice, locally Euclidean spaces. For example, consider the point **p** in the set in Figure 8.8(a) and the line **L** in the set in Figure 8.8(b). We can think of these as zero sets of the function $x^2 - y^2$ in the plane and 3-space, respectively. In the first case we cannot find any neighborhood of **p** that looks like a line. No matter how small the neighborhood, it will always look like a cross. A similar problem exists in the second case. It is possible to give examples where worse problems arise. This leads to the following question: Is there a simple criterion that can be applied to

f that guarantees that V(f) will be a nice space? First of all, note that this is a **local** question that has to do with what neighborhoods of points look like.

**8.3.4. Theorem.** (The Implicit Parameterization Theorem) Let $f: \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable map. If $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ and if the $m \times n$ Jacobian matrix $f' = (\partial f_i/\partial x_j)$ has constant rank $k < n$ on an open set containing V(f), then V(f) is a differentiable manifold in $\mathbf{R}^n$ of dimension $n - k$.

**Proof.** We need to show that we can parameterize a neighborhood of an arbitrary point $\mathbf{x}_0$ in V(f). Without loss of generality assume that $\mathbf{x}_0$ is the origin. By Theorem 4.4.5 we can get local diffeomorphisms g and h of neighborhoods of the origin in $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, so that

$$h(f(g(x_1, \ldots, x_n))) = (x_1, \ldots, x_k, 0, \ldots, 0).$$

The map g can now be used to define the parameterization of a neighborhood of $\mathbf{x}_0$ that we need.

**8.3.5. Example.** To analyze the set V(f) for the function $f(x,y,z) = z - x^2$.

**Solution.** We have that

$$f'(x, y, z) = (-2x, 0, 1).$$

Clearly, f′ has rank 1 on the zero set V(f) and so by Theorem 8.3.4 V(f) is a smooth surface (two-dimensional manifold). In fact, V(f) is the parabolic "trough" shown in Figure 8.9(a).

**8.3.6. Example.** To analyze the set V(f) for the function $f(x,y,z) = (x^2 + 3y^2 + 2z^2 - 1, z)$.

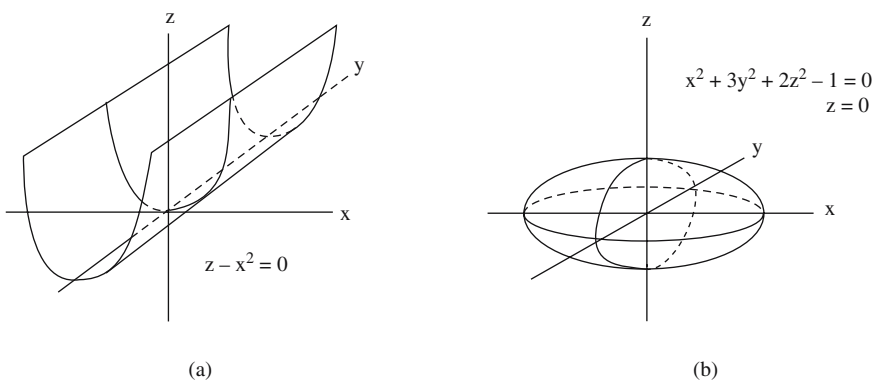



**Figure 8.9.** The varieties in Examples 8.3.5 and 8.3.6.

**Solution.** Since

$$f'(x,y,z) = \begin{pmatrix} 2x & 6y & 4z \\ 0 & 0 & 1 \end{pmatrix}$$

and f′ has rank 2 on the zero set V(f), Theorem 8.3.4 implies that V(f) is a smooth curve. This is easily verified because V(f), the set of common zeros of the functions

$$\begin{aligned} g(x,y,z) &= x^2 + 3y^2 + 2z^2 - 1 \\ h(x,y,z) &= z, \end{aligned}$$

is just the ellipse in the plane defined by the equation $x^2 + 3y^2 = 1$. See Figure 8.9(b).

The following theorem is a local version of Theorem 8.3.4.

**8.3.7. Theorem.** Let $f: \mathbf{R}^n \to \mathbf{R}^m$ and assume that $f(\mathbf{p}) = \mathbf{0}$. If the rank of Df is k in a neighborhood of $\mathbf{p}$, then there is a neighborhood $\mathbf{U}$ of $\mathbf{p}$ with $\mathbf{U} \cap \mathbf{V}(f)$ an $(n - k)$-dimensional manifold.

**Proof.** Since being a manifold is a local property, one can use the same argument as in Theorem 8.3.4.

Next we define what it means for a map between manifolds to be differentiable. It might seem as if there is nothing to do since Section 4.3 already defined a notion of differentiability for functions defined on subsets of $\mathbf{R}^m$. However, since our manifolds are not necessarily open subsets of $\mathbf{R}^m$, it is the definition given at the end of that section that would have to be used. Unfortunately, using that definition for the differentiability of a function on an arbitrary set one would not able to get a well-defined derivative of the function. Therefore, we shall use a definition based on the parameterizations of a manifold. After all, a parameterization corresponds to a coordinate system for a neighborhood of a point in the manifold and it makes sense to define differentiability with respect to such local coordinates.

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be $C^r$ manifolds in $\mathbf{R}^m$. A map $f: \mathbf{M}^n \to \mathbf{N}^k$ is said to be *of class* $C^r$ or a $C^r$ *map at a point* $\mathbf{p}$ in $\mathbf{M}^n$ if there is an open set $\mathbf{U}$ in $\mathbf{R}^n$, an open neighborhood $\mathbf{V}$ of $\mathbf{p}$ in $\mathbf{M}^n$, and a local $C^r$ parameterization $\Phi_{\mathbf{U},\mathbf{V}}: \mathbf{U} \to \mathbf{V}$, so that

$$f \circ \Phi_{\mathbf{U},\mathbf{V}}: \ \mathbf{U} \to \mathbf{N} \subseteq \mathbf{R}^m$$

is a $C^r$ map. The *rank of f at the point* $\mathbf{p}$ is the rank of $D(f \circ \Phi_{\mathbf{U},\mathbf{V}})$ at $\mathbf{u} = \Phi_{\mathbf{U},\mathbf{V}}^{-1}(\mathbf{p})$. The map f is a $C^r$ *map* if it is of class $C^r$ at every point $\mathbf{p}$ in $\mathbf{M}$. A *differentiable map* is a $C^\infty$ map.

See Figure 8.10. Notice that this definition does not yet define a derivative of the map f. We will do that in the next section because we need to define tangent vectors first. Right now we only have a notion of differentiability and rank. In this book we shall be mostly concerned with $C^\infty$ maps and not get involved in the fine points of $C^r$ maps, $n \neq \infty$.

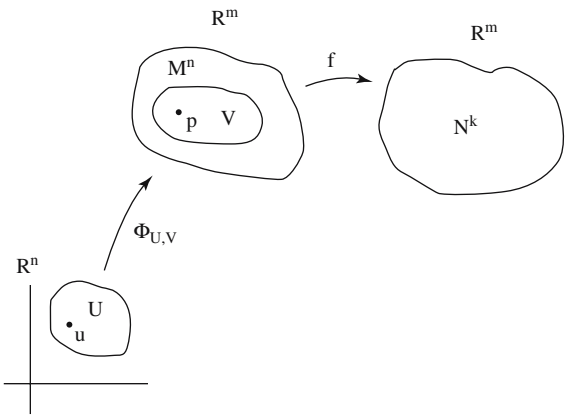



**Figure 8.10.** Defining a $C^r$ map between manifolds.

**8.3.8. Theorem.** The definition of differentiability and rank for maps between manifolds in Euclidean space is well defined and equivalent to the one given in Section 4.3.

**Proof.** To show that differentiability and rank are well defined we must show that the definitions do not depend on the local $C^r$ parameterization $\Phi_{\mathbf{U},\mathbf{V}}$. Let $\mathbf{U}'$ be another open set in $\mathbf{R}^n$, $\mathbf{V}'$ another open neighborhood of $\mathbf{p}$ in $\mathbf{M}^n$, and $\Phi_{\mathbf{U}',\mathbf{V}'}:\mathbf{U}' \to \mathbf{V}'$ another $C^r$ parameterization. The main observation is that

$$f \circ \Phi_{\mathbf{U}',\mathbf{V}'} = f \circ \Phi_{\mathbf{U},\mathbf{V}} \circ \left(\Phi_{\mathbf{U},\mathbf{V}}{}^{-1} \circ \Phi_{\mathbf{U}',\mathbf{V}'}\right)$$

and the map $\Phi_{\mathbf{U},\mathbf{V}}{}^{-1} \circ \Phi_{\mathbf{U}',\mathbf{V}'}$ is a $\mathbf{C}^R$ diffeomorphism between open sets in $\mathbf{R}^n$. We may have to restrict the maps to smaller open sets so that the composites are defined, but we shall leave the details to the reader. See also Theorem 4.3.25.

**8.3.9. Example.** To show that the antipodal map $f:\mathbf{S}^1 \to \mathbf{S}^1$ defined by $f(\mathbf{p}) = -\mathbf{p}$ is a $C^\infty$ map.

**Solution.** Let $\mathbf{p}_0 \in \mathbf{S}^1$ and assume that $\mathbf{p}_0 = (\cos\theta_0, \sin\theta_0)$. The map

$$\Phi\colon\ (\theta_0 - 1, \theta_0 + 1) \to \mathbf{S}^1$$

defined by

$$\Phi(\theta) = (\cos\theta, \sin\theta)$$

is a local $C^\infty$ parameterization of a neighborhood of $\mathbf{p}_0$. Now,

$$f \circ \Phi\colon\ (\theta_0 - 1, \theta_0 + 1) \to \mathbf{R}^2$$

is just

$$(f \circ \Phi)(\theta) = (-\cos\theta, -\sin\theta)$$

and this map is clearly $C^\infty$. If $\mathbf{p}_0 = (x_0, y_0)$ with $y_0 > 0$, then we could have chosen another local $C^\infty$ parameterization such as

$$\Psi\colon\ (-1,1) \to \mathbf{S}^1,$$

where

$$\Psi(x) = \left(x, \sqrt{1 - x^2}\right).$$

This time

$$(f \circ \Psi)(x) = \left(-x, -\sqrt{1 - x^2}\right),$$

which is also $C^\infty$.

**Definition.** Let $f\colon \mathbf{M}^n \to \mathbf{N}^k$ be a differentiable map between differentiable manifolds. If f has rank n at all points of $\mathbf{M}$, the f is said to be an *immersion*. If f is a homeomorphism onto $f(\mathbf{M}) \subseteq \mathbf{N}$ and is an immersion, then f is called an *imbedding*. If f is a homeomorphism between $\mathbf{M}$ and $\mathbf{N}$ and an immersion, then it is called a *diffeomorphism*.

Immersions may only be locally one-to-one. For example, a figure eight is an immersion of a circle in the plane but not an imbedding.

**8.3.10. Theorem.** If $f\colon \mathbf{M}^n \to \mathbf{N}^k$ is a diffeomorphism, then n = k and $f^{-1}\colon \mathbf{N} \to \mathbf{M}$ is also a diffeomorphism.

**Proof.** The fact that n = k follows from the invariance of domain theorem, Theorem 7.2.3.8. Since f is a homeomorphism, it has an inverse which is also a homeomorphism. To prove that $f^{-1}$ is a diffeomorphism, use the inverse function theorem, Theorem 4.4.2.

**Definition.** A differentiable manifold $\mathbf{N}^k$ that is a subset of a closed manifold $\mathbf{M}^n$ is called a (differentiable) *submanifold* of $\mathbf{M}^n$ if the inclusion map is an immersion. If the manifold $\mathbf{M}^n$ has a nonempty boundary, then we also require that for every $\mathbf{p} \in \mathbf{N}^k$, there is an open neighborhood $\mathbf{U}$ of $\mathbf{p}$ in $\mathbf{M}^n$, an imbedding $h\colon \mathbf{U} \to \mathbf{R}^n$, and an open subset $\mathbf{V} \subset \mathbf{R}^k{}_+ \subset \mathbf{R}^n$, so that $\mathbf{N}^k \cap \mathbf{U} = h^{-1}(\mathbf{V})$.

The reason for the complication in the case of manifolds with boundary is that we do not want $\mathbf{N}$ to meet the $\partial\mathbf{M}$ in a bad way. See Figure 8.11(a) for some cases of h, $\mathbf{U}$, and $\mathbf{V}$. Figure 8.11(b) shows some good submanifolds and Figure 8.11(c) some imbedded manifolds that we would not want to call submanifolds. Among other things, unless $\mathbf{N}$ is contained in $\partial\mathbf{M}$, $\mathbf{N} - \partial\mathbf{N}$ should not meet $\partial\mathbf{M}$ and $\partial\mathbf{N}$ should always meet $\partial\mathbf{M}$ nicely ("nice" means transversally as defined in Section 8.11).

Finally, another common term is the following:

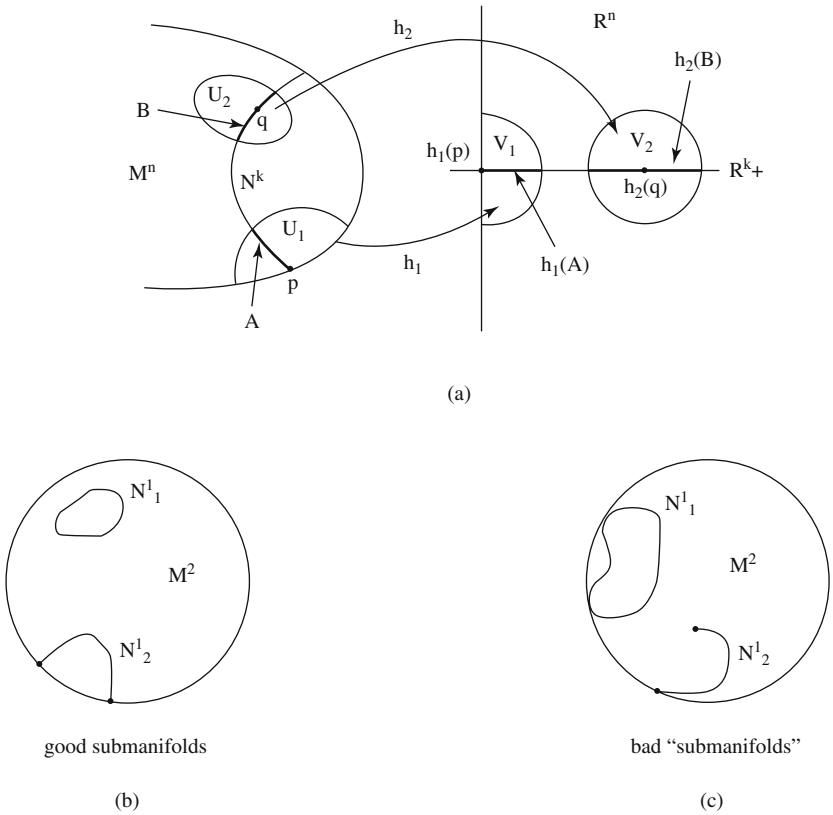



**Figure 8.11.** Good and bad "submanifolds".

**Definition.** A submanifold $\mathbf{N}^k$ of a manifold $\mathbf{M}^n$ is said to have *codimension* n – k in $\mathbf{M}^n$.

Unless stated otherwise, all manifolds from now on are assumed to be differentiable manifolds.

## 8.4 Tangent Vectors and Spaces

Curves are basic to understanding our definition of tangent vectors and tangent spaces of manifolds.

**Definition.** A $C^r$ *parametric curve* is a $C^r$ function $F:[a,b] \to \mathbf{R}^n$. The space $\mathbf{X} = F([a,b])$ traced out by F will be called the *path* of the parametric curve F. A *differentiable parametric curve* is a $C^\infty$ parametric curve. The parametric curve F is said to be

*closed* if F(a) = F(b). If F is closed and F is one-to-one on [a,b] or one-to-one on (a,b), then F will be called a *proper parametric curve* or *proper parameterization* of the path **X**.

Note that a parametric curve is a **function**. We have reserved the word "curve" by itself to mean a **set**, in the same way that "path" or "surface" refer to a set. The reader needs to be cautioned about the terminology. The problem is that these words are often used to mean either a set or a function with only the context making clear which is meant. Our terminology attempts to more or less allow for the common usage of the words dealing with curves and surfaces while at the same time maintaining the important distinction between a **set** and a **function**. As an example of this, we point out to the reader that expressions such as "differentiable parametric curve," "the parametric curve F(t)," etc., will often be abbreviated to "differentiable curve," "the curve F(t)," etc., respectively, in the future.

Let $F:[a,b] \to \mathbf{R}^n$ be a differentiable parametric curve. If

$$F(t) = (F_1(t), F_2(t), \ldots, F_n(t)),$$

then we know that

$$F'(t) = \left(F_1'(t), F_2'(t), \ldots, F_n'(t)\right).$$

**Definition.** The vector F′(t) is called the *tangent vector* of the curve F at F(t). In certain contexts it is called the *velocity* of F at F(t) and its length, |F′(t)|, is called the *speed* of F at F(t). The vectors

$$\frac{F'(t)}{|F'(t)|} \qquad (\text{or } \mathbf{0} \text{ if } F'(t) = \mathbf{0})$$

are called the *unit tangent vectors* of the curve.

Some simple cases and pictures should convince the reader that the term "tangent vector" makes sense because the vectors that one gets are indeed what one would want to call "tangent" to the curve. One can see this also from the definition of the derivative, which makes F′(t) a limit of secant lines. The terms "velocity" and "speed" also make sense because the derivative specifies a rate of change. Note that F′(t) is another curve and so one can keep differentiating (if the derivative exist) to get higher-order derivatives F″(t), . . . , $F^{(k)}(t)$, . . . of F(t).

**8.4.1. Example**. Define $F(t) = (t^2 + 3t, \sin t, 5)$. Then $F'(t) = (2t + 3, \cos t, 0)$ and the tangent vector to F at (0,0,5) is (3,1,0). Also, $F''(t) = (2, -\sin t, 0)$, and so F″(0) = (2,0,0).

**8.4.2. Example**. Consider the parameterization of the unit circle defined by $F(\theta) = (\cos\theta, \sin\theta)$. Then $F'(\theta) = (-\sin\theta, \cos\theta)$ and this vector is clearly a vector "tangent" to the circle at F(θ) since $F(\theta) \bullet F'(\theta) = 0$.

Next, note that if g(t) = t, then, using the chain rule, the tangent vector F′(t) is just

$$DF(t)(g'(t)) = DF(t)(1).$$

Thus, if we understand tangent vectors of the curve corresponding to the x-axis, then we get the tangent vector for an arbitrary parametric curve by mapping the tangent vectors along the x-axis over to it.

Now let us move on to higher dimensions. It clearly makes sense to talk about the tangent plane of a surface in $\mathbf{R}^3$ at a point. Higher-dimensional manifolds in $\mathbf{R}^n$ also have tangent planes. There are a number of different rigorous definitions that can be given for the tangent space of a submanifold of Euclidean space. The easiest way is to take advantage of the fact that we understand tangent vectors to parametric curves.

**Definition.** Let $\mathbf{M}$ be a manifold and let $\mathbf{p} \in \mathbf{M}$. If

$$\gamma\colon [-a,a] \to \mathbf{M}, a > 0,$$

is any parametric curve in $\mathbf{M}$ with $\gamma(0) = \mathbf{p}$, then $\gamma'(0)$ is called a *tangent vector of* $\boldsymbol{M}$ *at* $\boldsymbol{p}$. The set of tangent vectors to $\mathbf{M}$ at $\mathbf{p}$ is called the *tangent space of* $\boldsymbol{M}$ *at* $\boldsymbol{p}$ and is denoted by $T_{\mathbf{p}}$ or $T_{\mathbf{p}}(\mathbf{M})$. The set

$$\{\mathbf{p} + \mathbf{v} \mid \mathbf{v} \in T_{\mathbf{p}}\}$$

is called the *tangent plane of* $\boldsymbol{M}$ *at* $\boldsymbol{p}$.

See Figure 8.12.

**Caution.** Later on, such as in Sections 8.11, 8.12, and 9.16, we shall, for technical reasons, want to "index" tangent vectors by the point at which they are defined. Therefore the definition of "tangent vector" and "tangent space" will be changed slightly. This will in no way alter any facts about these concepts but only involve a simple and obvious translation of terminology. The later definitions are needed for the theory of abstract manifolds. Right now we are quite happy to stick to subspaces of $\mathbf{R}^n$ and to keep the notation as simple as possible.

Note that the tangent space and the tangent plane are just translations of each other. The tangent space passes through the origin (since the constant curve $\gamma(t) = \mathbf{p}$, produces the zero tangent vector) and the tangent plane at $\mathbf{p}$ passes through $\mathbf{p}$. To justify the terminology we need to show that the $T_{\mathbf{p}}(\mathbf{M})$ is a vector space. We also need a practical way to compute these spaces.

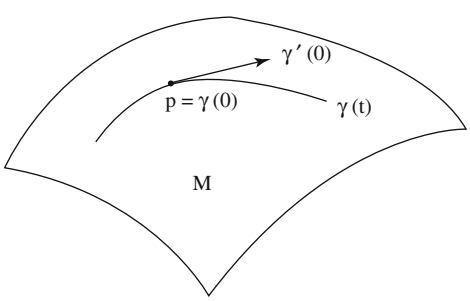



**Figure 8.12.** A tangent vector at $\mathbf{p}$ in a manifold $\mathbf{M}$.

**8.4.3. Theorem.** Let $\mathbf{M}$ be a k-dimensional submanifold of $\mathbf{R}^n$ and let $\mathbf{p} \in \mathbf{M}$. Let $\mathbf{V}$ be an open neighborhood of $\mathbf{p}$ in $\mathbf{M}$ and let $\Phi : \mathbf{U} \to \mathbf{V}$ be any parameterization of $\mathbf{V}$, where $\mathbf{U}$ is an open set in $\mathbf{R}^k$. Let $\mathbf{q} \in \mathbf{U}$ and $\Phi(\mathbf{q}) = \mathbf{p}$.

(1) If $\mathbf{M} = \mathbf{R}^k = \mathbf{R}^n$, then $T_{\mathbf{p}}(\mathbf{R}^k) = \mathbf{R}^k$.
(2) $T_{\mathbf{p}}(\mathbf{M})$ is a k-dimensional vector space. In fact, $T_{\mathbf{p}}(\mathbf{M}) = D\Phi(\mathbf{q})(\mathbf{R}^k)$.
(3) The tangent plane of $\mathbf{M}$ at $\mathbf{p}$ is a k-dimensional plane.
(4) The vectors $\dfrac{\partial \Phi}{\partial u_i}(\mathbf{q})$, $i = 1,2, \ldots, k$, are a basis for $T_{\mathbf{p}}(\mathbf{M})$.

**Proof.** Figure 8.13 should help the reader follow the proof. To prove (1) note that the lines through $\mathbf{q}$ parallel to the coordinate axes in $\mathbf{R}^k$ correspond to some very special curves $\gamma_i$ defined by

$$\gamma_i(t) = \mathbf{q} + t\mathbf{e}_i = (q_1, \ldots, q_{i-1}, q_i + t, q_{i+1}, \ldots, q_k).$$

Clearly, $\gamma_i(0) = \mathbf{q}$. Furthermore, it is also easy to check that $\gamma_i'(0) = \mathbf{e}_i$. This gives us a clue as to how to define a curve $\gamma(t)$ through $\mathbf{q}$ that has as a tangent vector an arbitrary vector $\mathbf{v} \in \mathbf{R}^k$. Simply let

$$\gamma(t) = \mathbf{q} + \sum_{i=1}^{k} t v_i \mathbf{e}_i = (q_1 + v_1 t, \ldots, q_k + v_k t).$$

Then $\gamma'(0) = \mathbf{v}$, from which (1) follows.

To prove (2) note that if $\gamma(t)$ is a parametric curve in $\mathbf{R}^k$ through $\mathbf{q}$, then $\mu(t) = \Phi(\gamma(t))$ is a curve in $\mathbf{M}$ through $\mathbf{p}$. The chain rule implies that

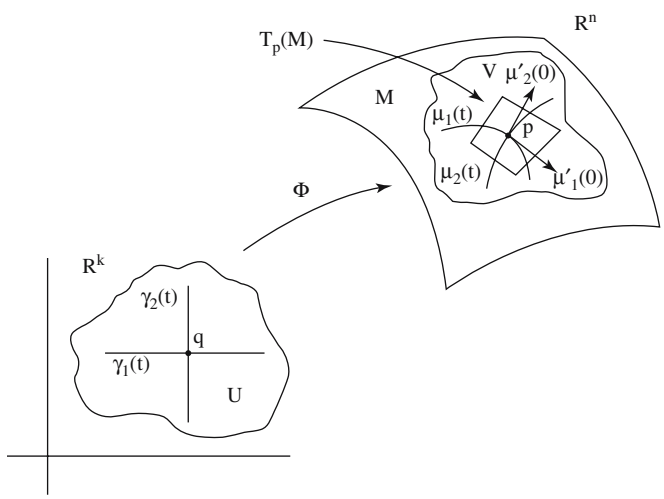



**Figure 8.13.** Defining the standard basis for the tangent space.

$$\mu'(0) = D\Phi(\mathbf{q})(\gamma'(0)). \tag{8.4}$$

This shows that $D\Phi$ maps tangent vectors to $\mathbf{R}^k$ at $\mathbf{q}$ to tangent vectors to $\mathbf{M}$ at $\mathbf{p}$, that is,

$$D\Phi(\mathbf{q}):\ T_{\mathbf{q}}(\mathbf{R}^k) = \mathbf{R}^k \to T_p(\mathbf{M}) \quad \left(\subseteq T_p(\mathbf{R}^n) = \mathbf{R}^n\right).$$

Since $\Phi$ is a local parameterization, $D\Phi(\mathbf{q})$ is a one-to-one linear map from $\mathbf{R}^k$ to $\mathbf{R}^n$. It remains to show that $D\Phi(\mathbf{q})$ is onto $T_{\mathbf{p}}(\mathbf{M})$.

Let $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{M})$ and let $\mu(t)$ be a curve in $\mathbf{M}$ with $\mu(0) = \mathbf{p}$ and $\mu'(0) = \mathbf{v}$. We need to find a curve $\gamma(t)$ in $\mathbf{R}^k$ with $\gamma(0) = \mathbf{q}$ and $\mu(t) = \Phi(\gamma(t))$, because then the chain rule (equation (8.4)) implies that

$$D\Phi(\mathbf{q})(\gamma'(0)) = \mathbf{v}.$$

But $\Phi^{-1}(\mu(t))$ is such a curve (there is no loss in generality in assuming that $\mu$ lies in $\mathbf{V}$) and (2) is proved.

Part (3) clearly follows from (2). Finally, we prove (4). Using the $\gamma_i$ defined above, define curves $\mu_i$ through $\mathbf{p}$ by $\mu_i(t) = \Phi(\gamma_i(t))$. The $\mu_i$ can be thought of as defining a local *curvilinear coordinate system* for $\mathbf{M}$ at $\mathbf{p}$. See Figure 8.13. Furthermore, the curves $\mu_i(t)$ are often called the *$u_i$-parameter curves* for the parameterization $\Phi(u_1, u_2, \ldots, u_k)$. In the special case of a surface, one would refer to $\mu_1(t)$ and $\mu_2(t)$ as the *u- and v-parameter curves* at $\mathbf{p}$ in $\mathbf{M}$, respectively.

Recalling how partial derivatives are defined it is easy to see that

$$\mu_i'(0) = \left(\frac{\partial \Phi_1}{\partial u_i}(\mathbf{q}), \ldots, \frac{\partial \Phi_n}{\partial u_i}(\mathbf{q})\right) = \frac{\partial \Phi}{\partial u_i}(\mathbf{q}).$$

On the other hand, the chain rule shown in equation (8.4) shows that

$$\mu_i'(0) = D\Phi(\mathbf{q})(\gamma_i'(0)) = D\Phi(\mathbf{q})(\mathbf{e}_i).$$

This proves (4).

**8.4.4. Corollary.** If $\mathbf{M}$ is a surface in $\mathbf{R}^3$, then the cross product $\partial\Phi/\partial x(\mathbf{q}) \times \partial\Phi/\partial y(\mathbf{q})$ is a normal vector for $T_{\mathbf{p}}(\mathbf{M})$.

**Note:** The fact that (4) holds in Theorem 8.4.3 was built into our definition of local parameterizations of a manifold because they are assumed to be regular. However, as mentioned earlier, it is sometimes natural to use parameterizations that are not regular. For example, consider the parameterization of the surface of revolution obtained by rotating the parabolic arc $y = 1 - x^2$, $-1 \le x \le 1$, about the x-axis. One can parameterize this surface via the map

$$\Phi(x, \theta) = \left(x, (1 - x^2)\cos\theta, (1 - x^2)\sin\theta\right), \quad -1 \le x \le 1,\ 0 \le \theta \le \pi.$$

Compare this with Example 8.2.2. This parameterization is differentiable everywhere but not regular and property (4) fails when x is ±1. Nonregularity may be a problem if one is interested in the tangent plane at those points where the parameterization is not regular because the standard way to find an equation for the tangent plane is to compute the normal using Corollary 8.4.4. Therefore, if condition (4) is important, then one needs to check that the parameterization at hand is in fact a regular parameterization at the points in question.

**8.4.5. Example.** To find the tangent plane **X** to $\mathbf{S}^2$ at the point $\mathbf{p} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$.

**Solution.** Let us use the parameterization

$$\Phi(u,v) = \left(u, v, \sqrt{1-u^2-v^2}\right)$$

for the upper hemisphere. Then

$$\frac{\partial \Phi}{\partial u} = \left(1, 0, -\frac{u}{\sqrt{1-u^2-v^2}}\right) \quad \text{and} \quad \frac{\partial \Phi}{\partial v} = \left(0, 1, -\frac{v}{\sqrt{1-u^2-v^2}}\right).$$

Since $\Phi\left(\frac{1}{\sqrt{2}}, 0\right) = \mathbf{p}$,

$$\frac{\partial \Phi}{\partial u}\left(\frac{1}{\sqrt{2}}, 0\right) = (1,0,-1) \quad \text{and} \quad \frac{\partial \Phi}{\partial v}\left(\frac{1}{\sqrt{2}}, 0\right) = (0,1,0)$$

are a basis for **X**. A normal vector for **X** is (1,0,1) = (1,0,–1) × (0,1,0). These answers clearly agree with one's intuition of what the plane should be.

Using a parameterization and Corollary 8.4.4 to determine the tangent plane involves a fair amount of computation. It turns out that if we are able to present our surface as the zeroes of a function, then it is much easier to get equations for the tangent planes. Compare the next result to Proposition 4.5.7.

**8.4.6. Proposition.** Let **M** be a manifold in $\mathbf{R}^n$ and suppose that $f: \mathbf{R}^n \to \mathbf{R}$ is a differentiable function such that **M** = V(f). Let $\mathbf{a} \in \mathbf{M}$. The tangent plane **X** of **M** at **a** is defined by the equation

$$\nabla f(\mathbf{a}) \bullet (\mathbf{p} - \mathbf{a}) = 0. \tag{8.5}$$

If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{p} = (x_1, x_2, \ldots, x_n)$, then this equation can be written as

$$\frac{\partial f}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{a})(x_n - a_n) = 0. \tag{8.6}$$

**Proof.** If $\gamma: [-c,c] \to \mathbf{M}$ is any curve in **M** through **a**, then

$$f(\gamma(t)) = 0$$

by hypothesis. Differentiating this equation using the chain rule yields

$$\nabla f(\mathbf{a}) \bullet \gamma'(0) = 0.$$

In other words, the gradient of f is orthogonal to every tangent vector. It follows that $\nabla f$ gives us a normal vector of the tangent plane that we are trying to find.

**8.4.7. Example.** To redo Example 8.4.5 using Propositions 8.4.6.

**Solution.** Define

$$f(x,y,z) = x^2 + y^2 + z^2 - 1.$$

The sphere is the zero set of f and $\nabla f = (2x,2y,2z)$. According to Proposition 8.4.6,

$$\nabla f\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = (\sqrt{2}, 0, \sqrt{2})$$

will be a normal vector to the plane. This agrees with the result in Example 8.4.5.

**8.4.8. Example.** To show that the lines normal to the tangent planes of an arbitrary sphere pass through its center.

**Solution.** Consider the sphere **S** of radius r with center $\mathbf{p}_0$. If

$$f(\mathbf{p}) = |\mathbf{p} - \mathbf{p}_0|^2 - r^2,$$

then $\mathbf{S} = V(f)$. It is easy to check that

$$\nabla f(\mathbf{p}) = 2(\mathbf{p} - \mathbf{p}_0).$$

Since the point-normal form of the line **L** through **p** and $\mathbf{p}_0$ is $\mathbf{p} + t(\mathbf{p} - \mathbf{p}_0)$, the normal $\nabla f(\mathbf{p})$ is a direction vector and we are done. See Figure 8.14.

Now we look at how differentiable maps induce natural maps on tangent spaces. Let

$$f\colon\ \mathbf{M}^n \to \mathbf{N}^k$$

be a differentiable map between differentiable manifolds. Let **p** be a point of $\mathbf{M}^n$ and let $\mathbf{q} = f(\mathbf{p})$. Define

$$Df_{\mathbf{p}}\colon\ \mathbf{T}_{\mathbf{p}}(\mathbf{M}^n) \to T_{\mathbf{q}}(\mathbf{N}^k)$$

as follows: Let **v** be any vector in the tangent space $T_{\mathbf{p}}(\mathbf{M}^n)$ and let $\gamma(t)$ be any curve lying $\mathbf{M}^n$ with $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$. Then

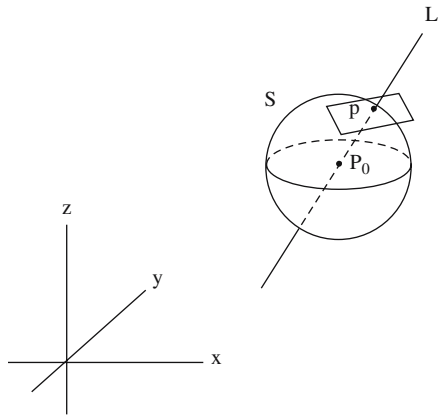



**Figure 8.14.** Normal lines to a sphere pass through its center.

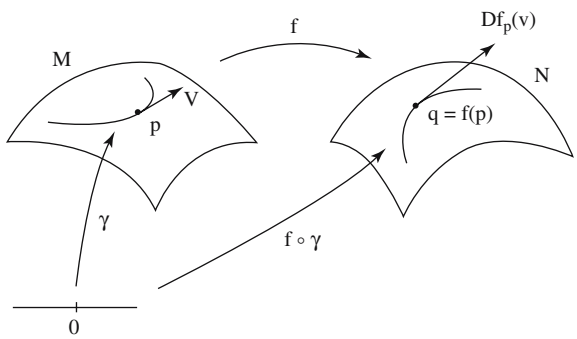



**Figure 8.15.** The derivative of a differentiable map between manifolds.

$$Df_{\mathbf{p}}(\mathbf{v}) = (f \circ \gamma)'(0). \tag{8.7}$$

See Figure 8.15.

**Definition.** The map $Df_{\mathbf{p}}$ is called the *derivative of f at the point* $\boldsymbol{p}$.

**8.4.9. Theorem.**

(1) The map $Df_{\mathbf{p}}$ is a well-defined linear map.
(2) If $\mathbf{M}^n = \mathbf{R}^n$ and $\mathbf{N}^k = \mathbf{R}^k$, then $Df_{\mathbf{p}} = Df(\mathbf{p})$.

**Proof.** There are two parts to the proof of (1). To show that the map is well defined one must show that its definition does not depend on the curve $\gamma$. In general, there will be lots of curves $\gamma(t)$ satisfying $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$. To show the linearity property, observe that

$$(f \circ \gamma)'(0) = D(f \circ \gamma)(0)(1),$$

and use the fact that $D(f \circ \gamma)(0)$ is a linear map. The details are left as Exercise 8.4.4.

Part (2) says that the map $Df_{\mathbf{p}}$ generalizes the ordinary definition of the derivative as defined in Section 4.3. Given a map $f: \mathbf{R}^n \to \mathbf{R}^k$, we defined a derivative $Df(\mathbf{p})$. If we think of $\mathbf{R}^n$ and $\mathbf{R}^k$ as differentiable manifolds, then we just defined a new derivative $Df_{\mathbf{p}}$. The two derivatives are the same linear maps because of the definitions involved and the fact that the tangent spaces at points of $\mathbf{R}^n$ and $\mathbf{R}^k$ are just $\mathbf{R}^n$ and $\mathbf{R}^k$, respectively (Theorem 8.4.3(2)).

**8.4.10. Example.** To compute $Df_x$ at a point $x \in \mathbf{R}$ for the map $f: \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^2$.

**Solution.** Let $\mathbf{v} \in T_x(\mathbf{R}) = \mathbf{R}$. The curve

$$\gamma\colon\ (-1,1) \to \mathbf{R},\ \ \gamma(t) = x + t\mathbf{v},$$

satisfies $\gamma(0) = x$ and $\gamma'(0) = \mathbf{v}$. But

$$(f \circ \gamma)(t) = (x + t\mathbf{v})^2,$$

so that

$$(f \circ \gamma)'(t) = 2(x + t\mathbf{v})\mathbf{v} \ \text{ and } \ (f \circ \gamma)'(0) = 2x\mathbf{v}.$$

We leave it to the reader to check that this agrees with $Df(x)(\mathbf{v})$. What would have happened if we had chosen the curve

$$\eta(t) = x + (t^3 + t)\mathbf{v},$$

which also satisfies $\eta(0) = x$ and $\eta'(0) = \mathbf{v}$? Well,

$$(f \circ \eta)(t) = [x + (t^3 + t)\mathbf{v}]^2$$

and

$$(f \circ \eta)'(t) = 2[x + (t^3 + t)\mathbf{v}](3t^2 + 1)\mathbf{v},$$

so that $(f \circ \eta)'(0) = 2x\mathbf{v}$, which agrees with the answer we got using the curve $\gamma(t)$.

## 8.5 Oriented Manifolds

This section returns to the topic of orientability. Section 1.6 looked at orientation in the context of vector spaces, which amounted to studying local orientations. In Section 7.5 we viewed orientation in the global context of (triangulated) pseudomanifolds. Now we want to describe orientation in the context of differentiable manifolds. The new definition will be compatible with the definition for pseudomanifolds, but will make use of the differential structure that we are assuming.

Let $\mathbf{M}^k$ be a k-dimensional submanifold of $\mathbf{R}^n$. Since each tangent space of $\mathbf{M}^k$ is a vector space we can talk about orientations in these tangent spaces.

**Definition.** Let $T:\mathbf{V} \to \mathbf{W}$ be an isomorphism between two k-dimensional vector spaces $\mathbf{V}$ and $\mathbf{W}$. Define

$$T_*: \text{ orientations of } \mathbf{V} \to \text{orientations of } \mathbf{W} \tag{8.8a}$$

by

$$\mathbf{T}_*([\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]) = [\mathbf{T}(\mathbf{v}_1), \mathbf{T}(\mathbf{v}_2), \ldots, \mathbf{T}(\mathbf{v}_k)], \tag{8.8b}$$

where $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an ordered basis of $\mathbf{V}$. If $\mu$ is an orientation of $\mathbf{V}$, then $T_*(\mu)$ is called the orientation of $\mathbf{W}$ *induced* by the isomorphism T.

It is easy to see that $T_*$ is a well-defined one-to-one correspondence between the orientations of $\mathbf{V}$ and $\mathbf{W}$. See Exercise 8.5.1.

**Definition.** Let $\sigma$ be a map that associates to each $\mathbf{p} \in \mathbf{M}^k$ an orientation of the tangent space $T_{\mathbf{p}}(\mathbf{M}^k)$. Such a choice is said to be a *continuously varying* choice of orientations if for every $\mathbf{p} \in \mathbf{M}^k$ there is an open neighborhood $\mathbf{V}$ of $\mathbf{p}$ in $\mathbf{M}^k$, a parameterization $\Phi:\mathbf{U} \to \mathbf{V}$ of $\mathbf{V}$ defined on an open set $\mathbf{U}$ in $\mathbf{R}^k$, and an orientation $\mu$ of $\mathbf{R}^k$ so that $D\Phi(\mathbf{q})_*(\mu) = \sigma(\Phi(\mathbf{q}))$, for $\mathbf{q} \in \mathbf{U}$.

The definition of continuously varying orientations is simpler than it may sound. See Figure 8.16. For example, in the case of a surface all it says is that if, say, the tangent plane at a point of the surface has been oriented in a "counter clockwise" fashion, then the tangent planes at nearby points are also oriented the same way. We want to exclude random choices of orientations—some counter clockwise and others clockwise. Note that we only chose one orientation $\mu$ of $\mathbf{R}^k$ and not one in each $T_{\mathbf{q}}(\mathbf{R}^k)$ because all those tangent spaces are the same and equal to $\mathbf{R}^k$ itself.

It is easy to show that the concept of continuously varying orientations does not depend on any particular parameterization. See Exercise 8.5.2.

**Definition.** An *orientation* of a differentiable manifold $\mathbf{M}$ is any continuously varying choice $\sigma$ of orientations for the tangent spaces of $\mathbf{M}$. A manifold is said to be *orientable* if it admits an orientation. An *oriented manifold* is a pair $(\mathbf{M},\sigma)$, where $\mathbf{M}$ is a manifold and $\sigma$ is an orientation for $\mathbf{M}$.

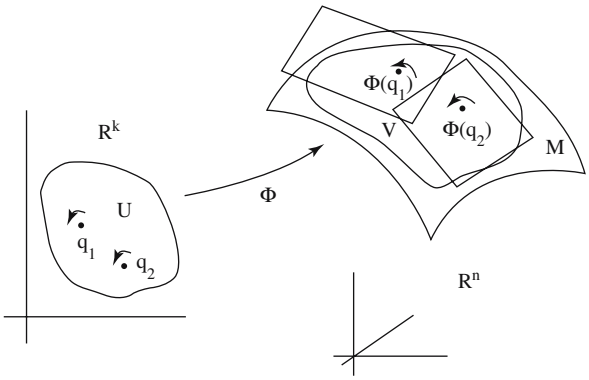



**Figure 8.16.** Continuously varying choice of orientation.

Exercise 8.5.3 describes another way to define the orientability of a manifold.

**Definition.** Let $\Phi: \mathbf{U} \to \mathbf{M}^k$ be a regular parameterization of a k-dimensional manifold $\mathbf{M}$. The orientation $\sigma$ of $\mathbf{M}$ that associates to each $\mathbf{p} = \Phi(\mathbf{q}) \in \mathbf{M}$ the orientation

$$\left[\frac{\partial \Phi}{\partial u_1}(\mathbf{q}), \frac{\partial \Phi}{\partial u_2}(\mathbf{q}), \ldots, \frac{\partial \Phi}{\partial u_k}(\mathbf{q})\right] \tag{8.9}$$

of the tangent space $T_{\mathbf{p}}(\mathbf{M})$ is called the *standard orientation of* $\boldsymbol{M}$ *induced by* $\boldsymbol{\Phi}$.

Dealing with ordered bases is not very convenient and so we now describe a better way to specify an orientation of a manifold in a common special case, but first some definitions.

**Definition.** Let $\mathbf{M}^k$ be a k-dimensional $C^r$ submanifold of $\mathbf{R}^n$. A *$C^r$ vector field* of $\mathbf{R}^n$ defined over $\mathbf{M}$ is a $C^r$ (vector-valued) function

$$\nu\colon\ \mathbf{M}^k \to \mathbf{R}^n.$$

The vector field $\nu$ is called *tangential to* $\boldsymbol{M}$ or simply a *$C^r$ vector field of* $\boldsymbol{M}$ if $\nu(\mathbf{p}) \in T_{\mathbf{p}}(\mathbf{M})$ for all $\mathbf{p} \in \mathbf{M}$. The vector field $\nu$ of $\mathbf{R}^n$ is called *normal to* $\boldsymbol{M}$ or a *$C^r$ normal vector field of* $\boldsymbol{M}$ *in* $\boldsymbol{R}^n$ if $\nu(\mathbf{p})$ is orthogonal to $T_{\mathbf{p}}(\mathbf{M})$ for all $\mathbf{p} \in \mathbf{M}$. (The phrase "in $\mathbf{R}^n$" is often dropped if $\mathbf{R}^n$ is clear from the context.) In any case we say that the vector field is a *unit vector field* or a *nonzero vector field* if $\nu(\mathbf{p})$ has unit length or is nonzero, respectively, for all $\mathbf{p} \in \mathbf{M}$.

Vector fields of manifolds associate vectors to points of a manifold with the vector at a point lying in the tangent space (or plane) at that point. Figure 8.17(a) shows a vector field of $\mathbf{S}^1$. Figure 8.17(b) shows a normal vector field of $\mathbf{S}^1$ in $\mathbf{R}^2$. As usual, the adjective "$C^r$" will be suppressed. The two typical cases are continuous ($C^0$) or $C^\infty$. Nonzero vector fields are often normalized to unit vector fields.

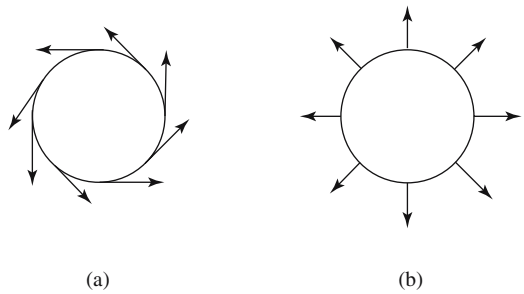



**Figure 8.17.** Tangent and normal vector fields on the circle.

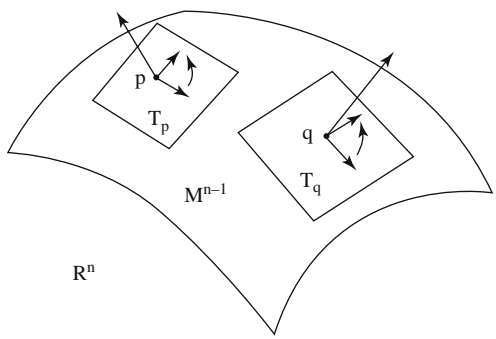



**Figure 8.18.** Normal vector fields and orientability.

With regard to the question of orientability, there is one case where normal vector fields are especially interesting.

**8.5.1. Theorem.** Let $\mathbf{M}^{n-1}$ be a submanifold of $\mathbf{R}^n$. Then $\mathbf{M}^{n-1}$ is orientable if and only if $\mathbf{M}^{n-1}$ admits a nonzero normal vector field.

**Proof.** See Figure 8.18. Since the tangent space at every point of $\mathbf{M}^{n-1}$ is an $(n-1)$-dimensional vector subspace of $\mathbf{R}^n$, for each point $\mathbf{p}$ of $\mathbf{M}^{n-1}$ we can express $\mathbf{R}^n$ uniquely as an orthogonal direct sum of the tangent space $T_{\mathbf{p}} = T_{\mathbf{p}}(\mathbf{M}^{n-1})$ and a one-dimensional subspace $N_{\mathbf{p}}$, that is,

$$\mathbf{R}^n = T_{\mathbf{p}} \oplus N_{\mathbf{p}}.$$

Suppose now that $\mathbf{M}^{n-1}$ is orientable. Then $\mathbf{M}^{n-1}$ admits a continuously varying choice $\sigma$ of orientations for its tangent spaces. Assume that $\sigma(\mathbf{p}) = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}]$ for some ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$ of $T_{\mathbf{p}}$. Choose that unique unit vector $\mathbf{n}_{\mathbf{p}}$ in $N_{\mathbf{p}}$ (there are two to choose from) so that the ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \mathbf{n}_{\mathbf{p}})$ induces the standard orientation of $\mathbf{R}^n$. It is not hard to show that the vector field

$$\nu\colon \mathbf{M}^{n-1} \to \mathbf{R}^n$$

defined by

$$\nu(\mathbf{p}) = \mathbf{n}_{\mathbf{p}} \tag{8.10}$$

is a nonzero normal vector field for $\mathbf{M}^{n-1}$.

Conversely, suppose that

$$\nu\colon \mathbf{M}^{n-1} \to \mathbf{R}^n$$

is a nonzero normal vector field for $\mathbf{M}^{n-1}$. For each $\mathbf{p} \in \mathbf{M}$, choose an ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$ of $T_{\mathbf{p}}$ so that $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \nu(\mathbf{p}))$ induces the standard orientation of $\mathbf{R}^n$. The map

$$\mathbf{p} \to [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}] \tag{8.11}$$

will define an orientation of $\mathbf{M}^{n-1}$.

It follows from Theorem 8.5.1, that we can define an orientation on an orientable $(n-1)$-dimensional manifold $\mathbf{M}^{n-1}$ in $\mathbf{R}^n$ simply by defining a nonzero normal vector field. This is typically the way one does it. Conversely, if a manifold is oriented, then the given orientation defines a unique unit normal vector field on it.

**8.5.2. Example.** The normal vector field

$$\nu\colon \mathbf{S}^2 \to \mathbf{R}^3$$

defined by $\nu(\mathbf{p}) = \mathbf{p}$ shows that the unit sphere $\mathbf{S}^2$ is orientable and defines the standard orientation of it.

**Definition.** Let $\Phi\colon \mathbf{U} \to \mathbf{M}^{n-1}$ be a regular parameterization of an $(n-1)$-dimensional submanifold $\mathbf{M}^{n-1}$ in $\mathbf{R}^n$. If $\sigma$ is an orientation of $\mathbf{M}^{n-1}$, then the unit normal vector field of $\mathbf{M}^{n-1}$ described in the proof of Theorem 8.5.1 and defined by equation (8.10) is called the *normal vector field of $\boldsymbol{M}^{n-1}$ in $\boldsymbol{R}^n$ induced by $\sigma$*. If $\sigma$ is the standard orientation of $\mathbf{M}^{n-1}$ induced by $\Phi$, then that normal vector field is called the *standard normal vector field of $\boldsymbol{M}^{n-1}$ induced by $\Phi$*.

**Note.** If an orientable submanifold $\mathbf{M}^{n-1}$ of $\mathbf{R}^n$ is closed (without boundary) and bounded, like for example the unit sphere in $\mathbf{R}^3$, then it divides space into bounded and unbounded parts and it makes sense to talk about "inward" and "outward" pointing normals for it.

Next, we address two related natural questions. First, we know that every compact connected differentiable manifold is a pseudomanifold (Theorem 8.3.3). Therefore, it is reasonable to ask whether the notion of orientable defined in this section is compatible with that given in Section 7.5. The second question is whether there are some simple criteria for determining the orientability of a manifold, since one certainly does

not want to define continuously varying families of orientations of tangent spaces or normal vector fields. To prove the compatibility of our two definitions of orientability one might be tempted to take a direct approach and show, for example, that orientations of tangent spaces induce coherent orientations of the simplices in a triangulation. Unfortunately, that would be technically complicated because homeomorphisms do not preserve any vector space structure. The better approach is to prove the next theorem from the definitions without reference to pseudomanifolds, using only properties of homology groups.

**8.5.3. Theorem.** A closed compact connected n-dimensional differentiable manifold $\mathbf{M}^n$ is orientable (according to the definitions in this section) if and only if $H_n(\mathbf{M}^n) \approx \mathbf{Z}$.

**Proof.** Conceptually the proof is not hard and consists of two steps:

**Step 1:** Relate an orientation of a tangent plane at a point to an orientation of a neighborhood of the point in $\mathbf{M}$.
**Step 2:** Relate a "continuously varying" collection of local orientations to a homology class.

Although we have not defined what is meant by the orientation of a neighborhood of a point it should be at least intuitively obvious. Think back to our discussion of the orientation of a surface. It is the old story. We have a concept in the linear setting of vector spaces and we extend it to curved spaces via a linearization process, that is, we use tangent planes to approximate the space locally. Filling in the details for Step 1 would be very messy given our current approach to manifolds. To do things more elegantly involves a more abstract approach that we cannot go into here. One would need to know about vector bundles (defined in Section 8.10) and more. More details for Step 1 can be found in [MilS74] or [Span66].

Step 2 is the more straightforward part and not that hard, but we are not able to present it here because it uses properties of homology groups we did not state or prove in Chapter 7. Roughly speaking, one shows that an orientation of $\mathbf{M}$ is equivalent to a choice of generator

$$\mu_{\mathbf{p}} \in H_n(\mathbf{M}, \mathbf{M} - \mathbf{p}) \approx \mathbf{Z}$$

for each $\mathbf{p} \in \mathbf{M}$ that "varies continuously" in the sense that for some compact neighborhood $\mathbf{N}$ of $\mathbf{p}$ there is a class

$$\mu_{\mathbf{N}} \in H_n(\mathbf{M}, \mathbf{M} - \mathbf{N})$$

that "restricts" to $\mu_{\mathbf{p}}$. The existence of such generators leads to a unique nonzero element

$$\mu_{\mathbf{M}} \in H_n(\mathbf{M}^n)$$

called the *fundamental homology class* of the oriented manifold that is a generator for $H_n(\mathbf{M}^n)$. Again see [MilS74] or [Span66].

**8.5.4. Corollary.** If $\mathbf{M}^n$ is a closed compact connected n-dimensional manifold, then it is orientable with the definition given in this section if and only if it is orientable with the definition of orientability of pseudomanifolds given in Section 7.5.

**Proof.** This corollary is an immediate consequence of Theorems 7.5.3, 8.3.3, and 8.5.3.

Using Theorem 8.5.3 it is now relatively easy to determine whether a connected differentiable manifold is orientable because we can compute the homology groups from the pseudomanifold structure.

We finish this section with a result about the existence of nonzero vector fields. Manifolds certainly admit nonzero (tangential) vector fields locally, that is, over sufficiently small neighborhoods of any given point. One gets this from the local parameterizations. On the other hand, it is an interesting question as to whether a manifold admits a global nonzero vector field. A famous result (Corollary 8.5.6 below) states that $\mathbf{S}^2$ does **not** have any nonzero vector field. This is the so called "hairy billiard ball" problem, which says that no matter how one combs a hairy billiard ball there will always be a discontinuity somewhere (a "cowlick"). See Figure 8.19. Of course, the "hairy circle" **can** be combed as Figure 8.17(a) shows. One can generalize the question about nonzero vector fields to asking how many linearly independent vector fields (the vectors are linearly independent at each point) a k-dimensional manifold admits. A beautiful deep result in topology answers this question completely. See [AgoM76]. What is beautiful about this is that it is a perfect example of where an answer to a question needs a good understanding of many fields in mathematics. It points out the virtue of having a broad knowledge and not just expertise in one specialty!

**8.5.5. Theorem.** $\mathbf{S}^n$ admits a nonzero vector field if and only if n is odd.

**Proof.** If n is odd, say n = 2k + 1, then

$$\sigma(x_1, x_2, \ldots, x_{2k+2}) = (-x_2, x_1, \ldots, -x_{2k+2}, x_{2k+1})$$

is a nonzero vector field. Next, suppose that n is even, say n = 2k, and that $\sigma$ is a nonzero vector field on $\mathbf{S}^n$. We may clearly assume that $\sigma(\mathbf{p})$ is a unit vector for all $\mathbf{p}$, so that we may consider $\sigma$ as a map from $\mathbf{S}^n$ to $\mathbf{S}^n$. Define

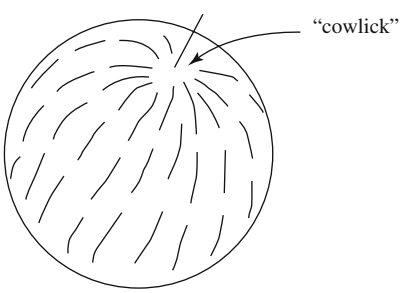



**Figure 8.19.** A "hairy" billiard ball cannot be combed.

$$h\colon \mathbf{S}^n \times [0,1] \to \mathbf{S}^n$$

by

$$h(\mathbf{p}, t) = (\cos \pi t)\mathbf{p} + (\sin \pi t)\sigma(\mathbf{p}).$$

If $C(\mathbf{p})$, $\mathbf{p} \in \mathbf{S}^n$, denotes the unit circle in the two-dimensional plane through the origin with orthonormal basis consisting of the vectors $\mathbf{p}$ and $\sigma(\mathbf{p})$, then $h(\mathbf{p},t)$ moves from $\mathbf{p}$ to $-\mathbf{p}$ on this circle $C(\mathbf{p})$ as t moves from 0 to 1. Thus, h is a homotopy between the identity map of $\mathbf{S}^n$ and the antipodal map. But the identity map has degree 1 and the antipodal map has degree $(-1)^{2k+1} = -1$ by Theorem 7.5.1.3. This contradiction shows that $\sigma$ cannot exist.

**8.5.6. Corollary.** An even-dimensional sphere, in particular $\mathbf{S}^2$, does not admit a nonzero vector field.

## 8.6 Handle Decompositions

This section begins the study of the topological structure of manifolds. We shall now make good on the promise we made in Section 4.6 to show how closely their structure is related to critical points on real-valued functions defined on them.

Before we get started, it is important that the reader understand certain notation used in this section. If an n-dimensional manifold $\mathbf{M}$ is a subset of $\mathbf{R}^m$, then the points of $\mathbf{M}$ are m-tuples, but this is the wrong way to look at them. Every point $\mathbf{p}$ of $\mathbf{M}$ has an open neighborhood $\mathbf{V}$ and a local parameterization $\varphi\colon \mathbf{U} \to \mathbf{V}$ defined on an open subset $\mathbf{U}$ of $\mathbf{R}^n$. The function $\varphi(u_1,u_2, \ldots ,u_n)$ defines a curvilinear coordinate system in $\mathbf{V}$, so that, typically

(1) one thinks of the point $\mathbf{p}$ as having coordinates $u_i$, and
(2) when one deals with a function f defined on $\mathbf{M}$ one thinks of f restricted to $\mathbf{V}$ as a function of the parameters $u_i$.

Formally, point (2) means that instead of working with the function f in a neighborhood of the point $\mathbf{p}$, one works with the function $f_\varphi(u_1,u_2, \ldots ,u_n) = f(\varphi(u_1,u_2, \ldots ,u_n))$. Because expressions would become cumbersome if one were to use the precise notation $f_\varphi$, one is sloppy and writes things like $f(u_1,u_2, \ldots ,u_n)$ in this case. There should be no confusion now that we have explained what is meant.

**Notation.** Expressions such as "*in local coordinates* $u_i$ the function f has the form $f(u_1,u_2, \ldots ,u_n) = \ldots$" will mean "$f_\varphi(u_1,u_2, \ldots ,u_n) = \ldots$"

This way of talking about functions on manifolds is very common. Of course, there are many local parameterizations $\varphi$ for $\mathbf{p}$, but it will not matter which we choose for what we want to do, so that we will not bother to mention $\varphi$ explicitly. (In other contexts, if things do depend on $\varphi$, then one has to take that dependence into account.)

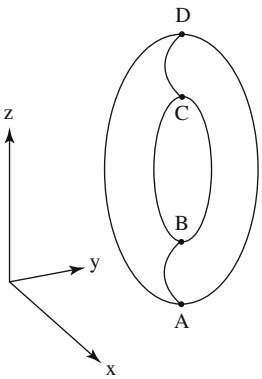



**Figure 8.20.** The height function for the torus.

In particular, when we talk about *critical points*, *critical values*, *nondegenerate critical points*, or the *index* of a critical point for functions $f: \mathbf{M} \to \mathbf{R}$ in this section, we shall mean these concepts as they were defined in Section 4.5 for $f_\varphi$. In Exercise 8.6.1 you are asked to show that everything is well defined and independent of the choice of $\varphi$.

Now let us return to the subject matter of this section. Let $\mathbf{M}$ be an n-dimensional closed manifold and consider a smooth real-valued function $f: \mathbf{M} \to \mathbf{R}$. Figure 8.20 shows the prototype of the kind of conclusions we want to draw. The figure shows a vertical torus and the function we have in mind is the height function f, that is,

$$f(\mathbf{p}) = \text{height of } \mathbf{p} \text{ above the x - y plane} = \text{z-coordinate of } \mathbf{p}.$$

The function f clearly has four critical points at $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ corresponding to critical values a, b, c, and d, respectively. Furthermore, these are non-degenerate critical points because in a neighborhood of these points the manifold looks like the graph of the functions

$$\begin{aligned} f_a(x,y) &= x^2 + y^2, \\ f_b(x,y) &= x^2 - y^2, \\ f_c(x,y) &= -x^2 + y^2, \quad \text{and} \\ f_d(x,y) &= -x^2 - y^2, \end{aligned}$$

respectively. Define

$$\mathbf{M}_s = f^{-1}((-\infty, s]).$$

How does $\mathbf{M}_s$ change as the value s changes? Well, $\mathbf{M}_s$ is the empty set if $s < 0$. If s is any number between 0 and a, then $\mathbf{M}_s$ is diffeomorphic to a disk. When s is between b and c, then $\mathbf{M}_s$ is diffeomorphic to a cylinder. When s is between c and d, then $\mathbf{M}_s$ is like a torus with a disk removed. Finally, $\mathbf{M}_s$ is the torus if $s \geq d$.

There is another way to describe the topological changes in $\mathbf{M}_s$. Consider the difference between $\mathbf{M}_{b-\varepsilon}$ and $\mathbf{M}_{b+\varepsilon}$. That change is equivalent to adding a "handle" to

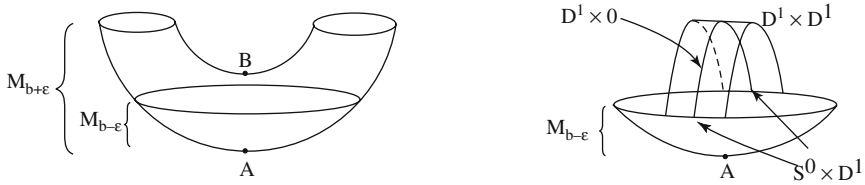



**Figure 8.21.** Passing a critical point adds a handle.

$\mathbf{M}_{b-\varepsilon}$. See Figure 8.21. By a handle we mean, in this case, that we have glued a rectangular strip $\mathbf{D}^1 \times \mathbf{D}^1$ to the boundary of $\mathbf{M}_{b-\varepsilon}$ along $\mathbf{D}^1 \times \mathbf{S}^0$. The "handle" part comes from $\mathbf{0} \times \mathbf{D}^1$. It is not accidental that the index of the critical point b is 1, which is also the "dimension" of the handle.

**8.6.1. Theorem.** Every closed compact differentiable manifold $\mathbf{M}$ admits a smooth real-valued function f which has only nondegenerate critical points with distinct critical values.

**Proof.** See [Miln65a].

A function f for $\mathbf{M}$ of the type guaranteed by Theorem 8.6.1 clearly has at least two critical points, namely, the two which correspond to the global minimum and maximum of f. It is also clear that it is possible to find an f that has an arbitrary number of nondegenerate critical points. We simply have to perturb f in a suitable nice way. This leads to some questions. What is the minimum number of critical points that f can have? For a fixed k, what can we say about the manifold if k is its minimum number of critical points? A sphere clearly admits an f that has precisely two critical points (at its north and south pole). If another manifold admits a function with only two critical points, is it diffeomorphic to a sphere? Is there any relation between the minimum number of critical points and algebraic invariants such as the homology groups? These questions will be addressed in the next section. Our first order of business is show what nondegenerate critical points imply about the local structure of a manifold.

**8.6.2. Theorem.** Let $f: \mathbf{M} \to \mathbf{R}$ be a smooth function with only nondegenerate critical points. Let $a < b$ and assume that $[a,b]$ does not contain any critical values of f and that

$$\mathbf{A} = f^{-1}([a,b])$$

is compact. Then $\mathbf{A}$ is diffeomorphic to $f^{-1}(a) \times [0,1]$. In particular, $\mathbf{M}_a$ is diffeomorphic to $\mathbf{M}_b$ and the inclusion map $\mathbf{M}_a \subset \mathbf{M}_b$ is a homotopy equivalence.

**Proof.** See [Miln63].

We may not omit the hypothesis that $\mathbf{A}$ is compact in Theorem 8.6.2.

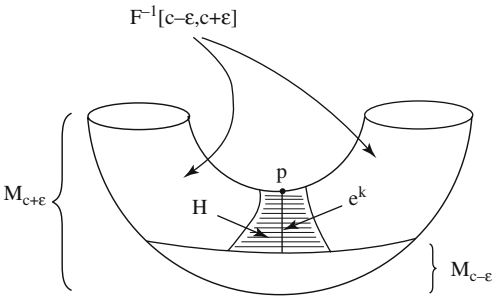



**Figure 8.22.** Deforming height function levels into handles.

**8.6.3. Theorem.** Let $f: \mathbf{M} \to \mathbf{R}$ be a smooth function and let $\mathbf{p}$ be a nondegenerate critical point for f with index k. Let $c = f(\mathbf{p})$. Assume that $f^{-1}([c - \varepsilon, c + \varepsilon])$ is compact and has no critical point other than $\mathbf{p}$ for some $\varepsilon > 0$. Then $\mathbf{M}_{c+\varepsilon}$ has the same homotopy type as $\mathbf{M}_{c-\varepsilon}$ with a k-cell attached.

**Proof.** We sketch a proof. For more details see [Miln63]. The height function for the vertical torus is again a good example for showing what we want to do. Consider Figure 8.22. The idea will be to deform f to a function $F: \mathbf{M} \to \mathbf{R}$ so that F is less that f in a small neighborhood of $\mathbf{p}$ and

$$\mathbf{A} = F^{-1}([-\infty, c - \varepsilon])$$

corresponds to $M_{c-\varepsilon}$ with the horizontally lined region labeled $\mathbf{H}$ attached. By pushing $\mathbf{H}$ in along the horizontal lines to the cell $e^k$ one shows that $\mathbf{A}$ has the homotopy type of $\mathbf{M}_{c-\varepsilon}$ with a k-cell attached. But $\mathbf{M}_{c+\varepsilon}$ can be contracted to $\mathbf{A}$ and so we will be done. We shall now fill in some of the details.

Now since $\mathbf{p}$ is a nondegenerate critical point, it follows from Theorem 4.6.3 that we can find local local coordinates $u_i$ for a neighborhood of $\mathbf{p}$ in which $\mathbf{p}$ corresponds to the origin $\mathbf{0}$ and the function f has the form

$$f(u_1, u_2, \ldots, u_n) = c - u_1^2 - \ldots - u_k^2 + u_{k+1}^2 + \ldots + u_n^2$$

in an open neighborhood $\mathbf{U}$ of the origin. The graph of f is easily analyzed in this coordinate system. Figure 8.23 tries to depict the general case. To understand the picture, try to imagine what one would see if one were to look vertically down at the critical point $\mathbf{p}$ in Figure 8.22

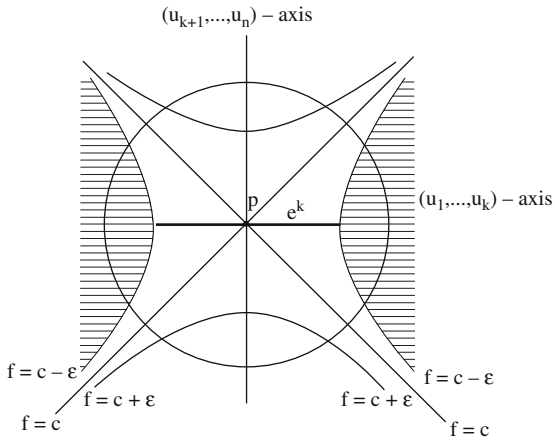
Choose a sufficiently small $\varepsilon > 0$, so that

(1) $f^{-1}([c - \varepsilon, c + \varepsilon])$ is compact and contains no critical point other than $\mathbf{p}$ and
(2) $\mathbf{U}$ contains the closed ball $\mathbf{B}$ of radius $2\varepsilon$ around $\mathbf{p}$.

Define

$$\mathbf{e}^k = \left\{(u_1, u_2, \ldots, u_n) \;\middle|\; u_1^2 + u_2^2 + \ldots u_k^2 \le \varepsilon \quad \text{and} \quad u_{k+1} = \ldots = u_n = 0\right\}.$$

**Figure 8.23.** Parameterizing a handle.




Clearly, $\partial \mathbf{e}^k = M_{c-\varepsilon} \cap \mathbf{e}^k$.

Next, define functions

$$\xi(u_1, u_2, \ldots, u_n) = u_1^2 + \ldots + u_k^2$$
$$\eta(u_1, u_2, \ldots u_n) = u_{k+1}^2 + \ldots + u_n^2,$$

so that

$$f = c - \xi + \eta,$$

and choose a $C^\infty$ function $\mu : \mathbf{R} \to \mathbf{R}$ satisfying

(1) $\mu(0) > \varepsilon$
(2) $\mu(x) = 0$ for $x \geq 2\varepsilon$, and
(3) $-1 < \mu'(x) \leq 0$ for all x.

Finally, using the local coordinates $u_i$, define the function

$$F\colon \mathbf{M} \to \mathbf{R}$$

in local coordinates by

$$\begin{aligned} F &= f - \mu(\xi + 2\eta) = c - \xi + \eta - \mu(\xi + 2\eta) && \text{on } \mathbf{U}, \\ &= f && \text{outside } \mathbf{U}. \end{aligned}$$

**Claim 1.** $F^{-1}([-\infty, c - \varepsilon]) = \mathbf{M}_{c+\varepsilon}$.

Claim 1 is proved by considering the ellipsoid defined by $\xi + 2\eta \leq 2\varepsilon$ and noting that the functions f and F agree outside the ellipsoid and inside it we have

$$\mathbf{F} \le f = c - \xi + \eta \le c + \frac{1}{2}\xi + \eta \le c + \varepsilon.$$

**Claim 2.** F and f have the same critical points.

To prove Claim 2, it suffices to prove that the only critical point of F inside **U** is **0**. But

$$\frac{\partial F}{\partial u_i} = \frac{\partial F}{\partial \xi}\frac{\partial \xi}{\partial u_i} + \frac{\partial F}{\partial \eta}\frac{\partial \eta}{\partial u_i}$$

and

$$\frac{\partial F}{\partial \xi} = -1 - \mu'(\xi + 2\eta) < 0$$
$$\frac{\partial F}{\partial \eta} = 1 - 2\mu'(\xi + 2\eta) \ge 1.$$

Since $\nabla\xi$ and $\nabla\eta$ only vanish at **0**, the same holds for $\nabla F$ and the claim is proved.

**Claim 3.** $F^{-1}([c-\varepsilon,c+\varepsilon])$ is compact and contains no critical points.

Claim 1 and the fact that $F \le f$ shows that

$$F^{-1}([c-\varepsilon,c+\varepsilon]) \subset f^{-1}([c-\varepsilon,c+\varepsilon]).$$

It follows that $F^{-1}([c-\varepsilon,c-\varepsilon])$ is compact. The only critical point it can contain is **0**, but this is impossible since

$$F(\mathbf{0}) = c - \mu(0) < c - \varepsilon.$$

Claim 3 is proved.

Next, define the region **H** by

$$\mathbf{H} = F^{-1}([-\infty, c-\varepsilon]) - \mathbf{M}_{c-\varepsilon}.$$

Then $F^{-1}([-\infty,c-\varepsilon]) = M_{c-\varepsilon} \cup \mathbf{H}$. Putting all these facts together proves Theorem 8.6.3.

**Definition.** The set $M_{c-\varepsilon} \cup \mathbf{H}$ is usually referred to as $M_{c-\varepsilon}$ with an *attached k-handle* ***H***.

**8.6.4. Theorem.** Let $f: \mathbf{M} \to \mathbf{R}$ be a smooth function that has only nondegenerate critical points. Assume that $\mathbf{M}_c$ is compact for all c. Then **M** has the homotopy type of a CW complex with one cell of dimension k for each critical point of f with index k.

**Proof.** See [Miln63].

We have just seen the close connection between the nondegenerate critical points of a real-valued function on a manifold and the manifold's topological structure. Here is another useful variant of some theorems that show this connection.

**8.6.5. Theorem.** (The Weak Morse Inequalities) Let **M** be a closed compact smooth n-dimensional manifold. Let $c_k$ be the number of critical points of index k for some function $f: \mathbf{M} \to \mathbf{R}$ which has only nondegenerate critical points. Let $\beta_k$ be the kth Betti number of **M**. Then

$$\beta_k \le c_k, \quad \text{and}$$
$$\chi(\mathbf{M}) = \sum_{k=0}^{n} (-1)^k \beta_k = \sum_{k=0}^{n} (-1)^k c_k.$$

**Proof.** See [Miln63]. This is an easy consequence of Theorem 8.6.4 and properties of the homology groups (Section 7.2.3).

We finish this section by summarizing its main results. Let **M** be a closed compact connected smooth n-dimensional manifold. We know that **M** admits a smooth function $f: \mathbf{M} \to \mathbf{R}$ that has only nondegenerate critical points. It is not hard to show that we may assume $f(\mathbf{M}) = [0,n]$ and that k is the critical value of all critical points of f with index k. With this assumption, we basically showed that **M** has (up to diffeomorphism) a filtration

$$\mathbf{D}^n = \mathbf{M}_0 \subset \mathbf{M}_1 \subset \ldots \subset \mathbf{M}_n = \mathbf{M},$$

where each $\mathbf{M}_k$ is obtained from $\mathbf{M}_{k-1}$ by attaching as many k-handles as there are critical points of index k. In other words,

$$\mathbf{M}_k = \mathbf{M}_{k-1} \cup \mathbf{D}^k \times \mathbf{D}^{n-k} \cup \mathbf{D}^k \times \mathbf{D}^{n-k} \cup \ldots \cup \mathbf{D}^k \times \mathbf{D}^{n-k},$$

where the attaching has taken place along $\mathbf{S}^{k-1} \times \mathbf{D}^{n-k}$ in each handle. Since **M** is connected, we may assume that there is only one handle of dimension 0 and n. This *handle decomposition* of a manifold is the starting point of the main classification results for manifolds. The reader should compare this with what we know about surfaces. Each connected closed surface can be constructed by starting with a disk and then adding a certain number of 1-handles and finally one disk to cap it off. The nonorientable surfaces are obtained by giving the 1-handles a "twist."

Again let **M** be a closed compact connected n-dimensional manifold and consider a smooth real-valued function

$$f: \mathbf{M} \to [0,n] \subset \mathbf{R}$$

with only nondegenerate critical points and so that 0 and n are the minimum and maximum values of f, respectively. We know that such functions f exist. This time rather than using f to build **M** using handles from the bottom up (thinking of f as a "height" function), let us build from the top down. To this end define

$$\mathbf{M}^s = f^{-1}([s,\infty]).$$

If we define

$$g = n - f\colon\ \mathbf{M} \to \mathbf{R},$$

then note that $\mathbf{M}^s$, as defined for f, is the same as $\mathbf{M}_{n-s}$, as defined for g.

Now let $\mathbf{p}$ be a critical point of index k for f and assume that it is the only critical point with critical value $c = f(\mathbf{p})$. Assume that f has no critical values in $[c - \varepsilon, c + \varepsilon]$ other than c. Let

$$\mathbf{A} = f^{-1}([c-\varepsilon, c+\varepsilon]).$$

This set is diffeomorphic to $\mathbf{D}^k \times \mathbf{D}^{n-k}$ and is the set that $\mathbf{M}^{c-\varepsilon}$ and $\mathbf{M}_{c+\varepsilon}$ have in common. It follows from what we did above that

$$\mathbf{M}^{c-\varepsilon} = \mathbf{M}^{c+\varepsilon} \cup_{\mathbf{D}^k \times \mathbf{S}^{n-k-1}} \mathbf{D}^k \times \mathbf{D}^{n-k}$$

and

$$\mathbf{M}_{c+\varepsilon} = \mathbf{M}_{c-\varepsilon} \cup_{\mathbf{S}^{k-1} \times \mathbf{D}^{n-k}} \mathbf{D}^k \times \mathbf{D}^{n-k}.$$

See Figure 8.24 and compare this to Figure 8.21. To put it another way, $\mathbf{M}^{c-\varepsilon}$ has the same homotopy type as $\mathbf{M}^{c+\varepsilon}$ with an n − k cell attached and $\mathbf{M}_{c+\varepsilon}$ has the homotopy type of $\mathbf{M}_{c-\varepsilon}$ with a k cell attached. Since we get the same space whether we start building it from the top or the bottom, what this shows is a fundamental duality between k and n − k cells. This is the Poincaré duality in Section 7.5.2.

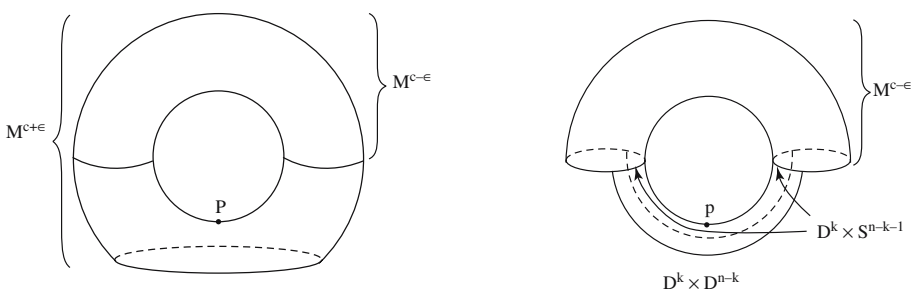



**Figure 8.24.** The dual handle decomposition.

## 8.7 Spherical Modifications

We begin this section by reinterpreting some of the results from the previous section. There we constructed manifolds by successively attaching a handle to the boundary of another manifold. The basic step had the form

$$\mathbf{N}' = \mathbf{N} \cup \mathbf{D}^k \times \mathbf{D}^{n-k}$$

where the handle $\mathbf{D}^k \times \mathbf{D}^{n-k}$ is attached to $\partial\mathbf{N}$ along $\mathbf{S}^{k-1} \times \mathbf{D}^{n-k}$. If we now concentrate on the boundaries of the manifolds, what has happened is that to get from $\partial\mathbf{N}$ to $\partial\mathbf{N}'$ we cut out $\mathbf{S}^{k-1} \times \mathbf{D}^{n-k}$ from $\partial\mathbf{N}$ and glued $\mathbf{D}^k \times S^{n-k-1}$ in its place along $\mathbf{S}^{k-1} \times \mathbf{S}^{n-k}$. We formalize this construction. Let $\mathbf{M}$ be a manifold and suppose that we have an imbedding

$$f\colon\ \mathbf{S}^{k-1} \times \mathbf{D}^{n-k} \to \mathbf{M}.$$

Define a manifold $\mathbf{M}'$ by

$$\mathbf{M}' = \left(\mathbf{M} - \operatorname{int}\left(\mathbf{S}^{k-1} \times \mathbf{D}^{n-k}\right)\right) \cup_{f|\mathbf{S}^{k-1}\times\mathbf{S}^{n-k-1}} \left(\mathbf{D}^k \times \mathbf{S}^{n-k-1}\right).$$

**Definition.** We shall say that the manifold $\mathbf{M}'$ is obtained from the manifold $\mathbf{M}$ by a *spherical modification of degree k* or by *surgery*.

**8.7.1. Theorem.** If $\mathbf{M}$ is obtained from $\mathbf{M}'$ via a spherical modification of degree k, then there is manifold $\mathbf{W}$ and a function $f\colon\mathbf{W} \to [0,1]$ satisfying

(1) $\partial\mathbf{W} = \mathbf{M} \cup \mathbf{M}'$,
(2) $f(\mathbf{M}) = 0$, $f(\mathbf{M}') = 1$, and
(3) f has a single nondegenerate critical point of index k in the interior of $\mathbf{W}$.

**Proof.** See [Wall68] or [Miln65a].

**Definition.** Two smooth n-dimensional manifolds $\mathbf{M}$ and $\mathbf{N}$ are said to be *cobordant* if there is a smooth (n + 1)-dimensional manifold $\mathbf{W}$ so that the boundary of $\mathbf{W}$ is the disjoint union of open and closed subsets $\mathbf{M}$ and $\mathbf{N}$. In that case, $\mathbf{W}$ is said to be a *cobordism* between $\mathbf{M}$ and $\mathbf{N}$.

**8.7.2. Theorem.** Two compact smooth n-dimensional manifolds are cobordant if and only if one can get from one to the other by a finite sequence of spherical modifications.

**Proof.** See [Wall68] or [Miln65a].

We state a few of the main results on the classification of manifolds most of which are consequences of the work of Smale [Smal61]. The key theorem is the next one.

**8.7.3. Theorem.** (The h-cobordism Theorem) Let $\mathbf{W}^n$ be a compact smooth n-dimensional manifold which is a cobordism between submanifolds $\mathbf{V}$ and $\mathbf{V}'$ satisfying the following conditions:

(1) All three spaces $\mathbf{W}$, $\mathbf{V}$, and $\mathbf{V}'$ are simply connected.
(2) The inclusion $\iota : \mathbf{V} \subset \mathbf{W}$ induces isomorphisms $\iota_{q*} : \mathbf{H}_q(\mathbf{V}) \to \mathbf{H}_q(\mathbf{W})$ for all q.
(3) $n \geq 6$.

Then $\mathbf{W}$ is diffeomorphic to $\mathbf{V} \times [0,1]$.

**Proof.** See [Miln65a].

**Definition.** Let $\mathbf{W}^n$ be a compact smooth n-dimensional manifold and which is a cobordism between submanifolds $\mathbf{V}$ and $\mathbf{V}'$. If both $\mathbf{V}$ and $\mathbf{V}'$ are deformation retracts of $\mathbf{W}$, then $\mathbf{W}$ is said to be an *h-cobordism* between $\mathbf{V}$ and $\mathbf{V}'$ and $\mathbf{V}$ and $\mathbf{V}'$ are said to be *h-cobordant*.

Theorem 8.7.3 gets its name from the fact that the hypotheses on $\mathbf{W}$, $\mathbf{V}$, and $\mathbf{V}'$ made $\mathbf{W}$ into an h-cobordism. (One needs Theorem 7.4.3.7 to see this.) There are some important corollaries.

**8.7.4. Corollary.** Two simply connected closed smooth n-dimensional manifolds, $n \geq 5$, which are h-cobordant are diffeomorphic.

The next corollary provides a characterization of the n-disk.

**8.7.5. Corollary.** Let $\mathbf{W}^n$ be a compact smooth simply connected n-dimensional manifold, $n \geq 6$, with simply connected boundary. Then the following four assertions are equivalent:

(1) $\mathbf{W}^n$ is diffeomorphic to $\mathbf{D}^n$.
(2) $\mathbf{W}^n$ is homeomorphic to $\mathbf{D}^n$.
(3) $\mathbf{W}^n$ is contractible.
(4) $\mathbf{W}^n$ has the same homology groups as a point.

In 1904 Poincaré conjectured the following:

**The Poincaré Conjecture:** Every compact simply connected closed three-dimensional manifold is homeomorphic to $\mathbf{S}^3$.

A generalization of this conjecture can be proved.

**8.7.6. Corollary.** (The Generalized Poincaré Conjecture) If $\mathbf{M}^n$, $n \geq 4$, is a closed compact simply connected smooth manifold which has the same homology groups as the n-sphere $\mathbf{S}^n$, then $\mathbf{M}^n$ is homeomorphic to $\mathbf{S}^n$.

**Proof.** When $n \geq 5$, then the result is an easy consequence of Corollary 8.7.5 and Theorem 7.5.2.7. The case $n = 4$ is more difficult and was proved in [Free82]. Note

that because of Poincaré duality we really only need to assume that the ith homology groups vanish for $0 < i \leq n/2$. The hypothesis of the manifold being simply connected is essential.

Corollary 8.7.6 cannot be strengthened to diffeomorphic because it is known, for example, that there are seven-dimensional manifolds $\mathbf{M}^7$ that are homeomorphic but **not** diffeomorphic to $\mathbf{S}^7$. The case $n = 2$ is trivially true, given what we know about surfaces. The original Poincaré conjecture for dimension 3 is unproved to this day, although at the time of this writing a proof may finally have been found (see [Miln03], which also contains a nice overview of the history of the conjecture). The main reason that the proofs in higher dimensions do not work in dimension 3 is that they rely on one's ability to deform continuous maps into imbeddings and in three dimensions we do not have enough space to be able to prove that this can always be done. It should be noted that an algorithm that determines whether or not a triangulated, closed, orientable three-dimensional manifold is homeomorphic to the 3-sphere **does** exist. See [Thom98] for a discussion of this algorithm that was discovered by J.H. Rubinstein in 1992. The proof depends on piecewise linear minimal surface theory. Unfortunately, it is exponential in the number of tetrahedra and not practically useful, even in the case of a very small number of tetrahedra.

Finally, we return to the problem of finding minimal cell decompositions for a manifold. This problem is equivalent to finding a minimal handle decomposition. Although we know that a manifold admits a handle decomposition, we may not get a minimal one. For example, Figure 8.25 shows how the height function applied to that sphere would produce a handle decomposition for the sphere which has one 0-handle, one 1-handle, and two 2-handles. The basic idea for finding minimal handle decompositions is to start with any decomposition and then to simplify it by letting handles cancel each other. In Figure 8.25, the 1-handle and one of the 2-handles cancel each other leaving a single 0- and 2-handle, which is a minimal handle decomposition for the sphere.

**8.7.7. Theorem.** (The Minimum Handle Decomposition Theorem) If $\mathbf{M}^n$, $n > 5$, is a simply connected compact closed differentiable manifold, then there is a nondegenerate $C^\infty$ function f with a minimum number of critical points consistent with the homology structure. We may assume that the value of the function at its critical points

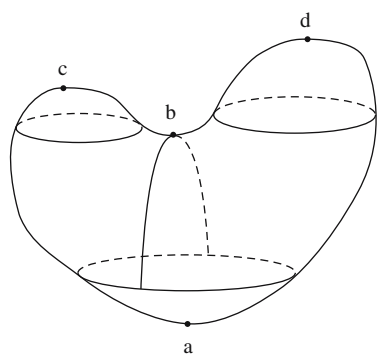



**Figure 8.25.** A nonminimal handle decomposition for a sphere.

is its index. More precisely, if $H_i(\mathbf{M})$ has rank r(i) and its torsion subgroup is a direct sum of q(i) cyclic groups, then we may assume that f has p(i) + q(i) + q(i − 1) critical points of index i.

**Proof.** See [Smal62].

There is a corresponding version of Theorem 8.7.7 for manifolds with boundary that asserts that we can find a function relative to the boundary with a minimum number of critical points.

Here are some final comments on the classification of manifolds. Clearly, arbitrary spaces cannot be classified up to homeomorphism using the homology, cohomology, or homotopy groups. However, in the case of compact closed manifolds one might have more hope along these lines because of their homogeneous structure and theorems like the Poincaré duality theorem. It turns out that two compact closed surfaces that are not homeomorphic do not have the same homotopy type. This follows from the fact that if the surfaces are not homeomorphic then they have nonisomorphic fundamental groups. See [Mass67]. On the other hand, such a theorem is false for 3-manifolds. Specifically, we pointed out in Example 7.2.4.7 that there are lens spaces that have the same homotopy type but are not homeomorphic.

## 8.8 Abstract Manifolds

Our definition of a manifold as a subset of $\mathbf{R}^n$ in Section 8.3 was not an intrinsic definition. This section sketches how to define manifolds in an intrinsic way. We need to get rid of the surrounding space. Rather than saying when a subset of Euclidean space is a manifold we shall say when an arbitrary topological space is a manifold. In that way we shall capture the intrinsic properties of such a space and not be distracted by how or where they are imbedded and other special properties of $\mathbf{R}^n$. For example, our universe is generally conceived as a three-dimensional manifold (or higher dimensional if one includes time or other dimensions). We do not think of our universe as imbedded in another space but rather as being all there is.

To follow the next definitions see Figure 8.26.

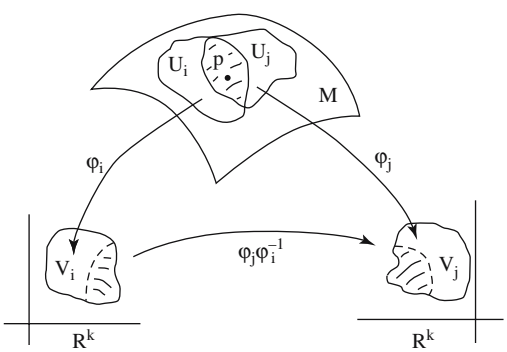



**Figure 8.26.** Defining a manifold intrinsically.

**Definition.** A (k-dimensional) *coordinate neighborhood* for a topological space **M** is a pair $(\mathbf{U},\varphi)$, where **U** is an open subset of **M** and $\varphi:\mathbf{U}\to\mathbf{V}$ is a homeomorphism onto an open subset **V** of $\mathbf{R}^k_+$.

Note that the map $\varphi$ for a coordinate neighborhood goes in the opposite direction from that of a local parameterization for a submanifold of $\mathbf{R}^n$ as defined in Section 8.3. Compare Figure 8.6 with Figure 8.26.

**Definition.** Let $r \geq 1$. A *$C^r$ differentiable structure* or *$C^r$ structure* for a k-dimensional topological manifold **M** is an indexed collection $S = \{(\mathbf{U}_i,\varphi_i)\}$ of k-dimensional coordinate neighborhoods for **M** satisfying the following conditions:

(1) The sets $\mathbf{U}_i$ cover **M**.
(2) If $\mathbf{U}_{ij} = \mathbf{U}_i \cap \mathbf{U}_j \neq \phi$, then the homeomorphism

$$\varphi_{ji} = \varphi_j \circ \varphi_i^{-1} : \varphi_i(\mathbf{U}_{ij}) \to \varphi_j(\mathbf{U}_{ij})$$

is a $C^r$ map.
(3) The collection S is maximal with respect to condition (2), that is, adding any other coordinate neighborhood $(\mathbf{U},\varphi)$ to S would violate that condition.

A *$C^r$ manifold* is a topological manifold **M** together with a $C^r$ differentiable structure for it. A *$C^0$ manifold* will simply mean a topological manifold. A $C^\infty$ manifold is called simply a *differentiable* or *smooth manifold*.

**8.8.1. Theorem.** Let $\mathbf{M}^k$ be an k-dimensional topological manifold.

(1) Any collection of coordinate neighborhoods for **M** satisfying (1) and (2) will always extend to a unique collection satisfying (3).
(2) Any $C^r$ structure of **M** induces a well-defined $(k-1)$-dimensional $C^r$ structure on $\partial\mathbf{M}$ making it into a $C^r$ manifold without boundary by using the restrictions of the relevant coordinate neighborhoods of **M** to the boundary.

**Proof.** This is an easy exercise.

**Note.** Condition (3) in the definition of a $C^r$ structure is a technical condition to give us freedom in choosing coordinate neighborhoods. Because of Theorem 8.8.1(1), to define a $C^r$ structure on a manifold, all that one ever bothers to do is define a collection of coordinate neighborhoods that satisfy conditions (1) and (2). Because of condition (3), the indices i and j in the definition belong in general to some uncountable set and are not intended to connote integers.

**8.8.2. Example.** The coordinate neighborhood ($\mathbf{R}^n$, identity map) induces a $C^\infty$ structure on $\mathbf{R}^n$, called the *standard $C^\infty$ structure*, making it into an n-dimensional $C^\infty$ manifold.

**8.8.3. Example.** The coordinate neighborhoods $(\mathbf{U}_i,\varphi_i)$, $i = 1,2,3,4$, where

$$\varphi_i : \mathbf{U}_i \to (-1,1)$$

and

$$\begin{aligned}
\mathbf{U}_1 &= \mathbf{S}^1{}_+, & \varphi_1(x,y) &= x, \\
\mathbf{U}_2 &= \{(x,y) \in \mathbf{S}^1 \mid x \geq 0\}, & \varphi_2(x,y) &= y, \\
\mathbf{U}_3 &= \mathbf{S}^1{}_-, & \varphi_3(x,y) &= x, \\
\mathbf{U}_4 &= \{(x,y) \in \mathbf{S}^1 \mid x \leq 0\}, & \varphi_4(x,y) &= y,
\end{aligned}$$

induce a $C^\infty$ structure on the unit circle $\mathbf{S}^1$ (Exercise 8.8.1(a)). Note that the maps $\varphi_i$ are just the inverses of the maps $\Phi_i$ in Example 8.3.2.

**8.8.4. Example.** To find a $C^\infty$ structure for the unit sphere $\mathbf{S}^n$, $n \geq 1$.

**Solution.** Rather than generalizing the coordinate neighborhoods defined in Example 8.8.3, we describe another collection $\{(\mathbf{U}_+,\varphi_+),(\mathbf{U}_-,\varphi_-)\}$:

$$\begin{aligned}
\mathbf{U}_+ &= \mathbf{S}^n - \{\mathbf{e}_{n+1}\}, & \varphi_+ &: \mathbf{U}_+ \to \mathbf{R}^n, \\
\mathbf{U}_- &= \mathbf{S}^n - \{-\mathbf{e}_{n+1}\}, & \varphi_- &: \mathbf{U}_- \to \mathbf{R}^n,
\end{aligned}$$

where $\varphi_+$ and $\varphi_-$ are the stereographic projections from $\mathbf{e}_{n+1}$ and $-\mathbf{e}_{n+1}$, respectively, that is, $\varphi_+ = p_n$ and $\varphi_- = p_n \circ r$, where r is the reflection of $\mathbf{R}^{n+1}$ about the plane $x_{n+1} = 0$. These coordinate neighborhoods induce a $C^\infty$ structure on $\mathbf{S}^n$ (Exercise 8.8.1(b)).

Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be $C^r$ manifolds with $C^r$ structure coordinate neighborhoods $\{(\mathbf{U}_i,\varphi_i)\}$ and $\{(\mathbf{V}_j,\psi_j)\}$, respectively. Define maps

$$\eta_{ij} : \mathbf{U}_i \times \mathbf{V}_j \to \mathbf{R}^n \times \mathbf{R}^k = \mathbf{R}^{n+k}$$

by

$$\eta_{ij}(\mathbf{u},\mathbf{v}) = (\varphi_i(\mathbf{u}), \psi_j(\mathbf{v})).$$

It is easy to check that the coordinate neighborhoods $\{(\mathbf{U}_i \times \mathbf{V}_j,\eta_{ij})\}$ induce a $C^r$ structure on $\mathbf{M} \times \mathbf{N}$ unless both manifolds have boundary, in which case one has to make some modifications to get legitimate coordinate neighborhoods for the points of $\partial\mathbf{M} \times \partial\mathbf{N}$. We will not present the details here, but they are straightforward.

**Definition.** The $C^r$ structure induced by the coordinate neighborhoods $\{(\mathbf{U}_i \times \mathbf{V}_j,\eta_{ij})\}$ is called the *product $C^r$ structure*. It makes $\mathbf{M} \times \mathbf{N}$ into an $(n + k)$-dimensional $C^r$ manifold called the *product $C^r$ manifold*.

Next, we define the concept of a differentiable map.

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be $C^r$ manifolds with coordinate neighborhoods $\{(\mathbf{U}_i,\varphi_i)\}$ and $\{(\mathbf{V}_j,\psi_j)\}$, respectively. A map $f : \mathbf{M}^n \to \mathbf{N}^k$ is said to be of *class $C^r$* or a *$C^r$ map* if

$$\psi_j \circ f \circ \varphi_i^{-1} : \varphi_i(\mathbf{U}_i) \to \mathbf{R}^k$$

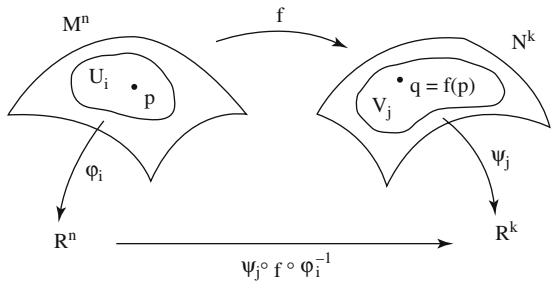



**Figure 8.27.** Differentiability of maps between manifolds.

is a $C^r$ map for all i and j. A $C^\infty$ map is called a *differentiable* map. If $r \geq 1$, then the *rank of f at a point* **p** of $\mathbf{M}^n$ is defined to be the rank of $D(\psi_j \circ f \circ \varphi_i^{-1})$ at **p** where $\mathbf{U}_i$ is a neighborhood of **p** and $\mathbf{V}_j$ is a neighborhood of f(**p**). If $\mathbf{A} \subseteq \mathbf{M}^n$, then the set of $C^r$ maps $f: \mathbf{A} \to \mathbf{R}$ will be denoted by $C^r(\mathbf{A})$.

See Figure 8.27. Differentiability and rank are well-defined local properties. One can show that a map is differentiable if and only if for each **p** in **M** and some $\mathbf{U}_i$ and $\mathbf{V}_j$ that contain **p** and f(**p**) in their interior, respectively, the map $\psi_j \circ f \circ \varphi_i^{-1}$ is differentiable. In other words, if one has differentiability with respect one pair of neighborhoods one has it for all. Furthermore, once one has a notion of a $C^r$ map and its rank, the terms of *immersion*, *imbedding*, *diffeomorphism*, and *submanifold* are defined as before and we do not duplicate the definitions here.

Manifolds admit different differentiable structures, but just because they are different does not mean that the resulting differentiable manifolds are not diffeomorphic. Calling two different differentiable structures on a manifold to be *equivalent* if they define diffeomorphic differentiable manifolds defines an equivalence relation on the set of differentiable structures. A famous problem that was largely solved in the 1960s and 1970s was to determine how many nonequivalent differentiable structures a manifold admitted. The surprising answer is that it is more than one in general. For example, the comment after Corollary 8.7.6 about the seven-dimensional sphere $\mathbf{S}^7$ can be interpreted as saying that $\mathbf{S}^7$ admits nonequivalent differentiable structures. In fact, one can show (see [FreL89]) that there are "fake" $\mathbf{R}^4$s, that is, four-dimensional manifolds that are homeomorphic to $\mathbf{R}^4$ but not diffeomorphic to it, so that even as simple a space as $\mathbf{R}^4$ admits nonequivalent differentiable structures. (On the other hand, this is also an example that the fourth dimension is special, because there are no fake $\mathbf{R}^n$s for $n \neq 4$.)

**Note.** In light of the fact that a manifold can have different differentiable structures, what happens in the case of some of the well-known spaces such as $\mathbf{R}^n$, $\mathbf{S}^n$, a torus, etc.? When we treat them as differentiable manifolds, which differentiable structure are we talking about? Well, these spaces invariably have some fairly "obvious" coordinate neighborhoods associated to them and so one assumes that these are being used for the differentiable structure and one never mentions differentiable structures explicitly. Although there may be more than one "obvious" collection of coordinate neighborhoods, they will lead to the same or diffeomorphic differentiable manifolds.

The really different differentiable structures are rather exotic. Also, when one builds new manifolds from old ones, the constructions invariably induce a natural differentiable structure on the new manifolds from the differentiable structure on the old ones. Finally and most importantly, the actual differentiable structure usually does not matter, only that one has one.

The reader may be wondering how differentiable manifolds differ from topological manifolds. A deep result in the theory of manifolds says that not all topological manifolds admit a differentiable structure. Now, the boundary of a square is a topological manifold that is not a differentiable submanifold of $\mathbf{R}^n$, but being homeomorphic to $\mathbf{S}^1$, it **does** admit a nice differentiable structure. Therefore, saying that a manifold does not admit a differentiable structure is saying much more, namely, that **every** imbedding of it in $\mathbf{R}^n$ has "corners." It follows that there is a big difference between $C^0$ and $C^1$ manifolds, but it turns out that there is essentially no difference between $C^1$ and $C^\infty$ manifolds (see [Munk61] or [Hirs76]). For this reason and in order not to get bogged down with technical issues as to how much differentiability one needs for a result, we mostly consider $C^\infty$ manifolds. Also, when it comes to maps, the next theorem shows that we can basically assume that all maps between differentiable manifolds are differentiable.

**8.8.5. Theorem.** Any continuous map between differentiable manifolds can be "approximated" by a differentiable map that is homotopic to the original. Furthermore, any two homotopic differentiable maps are homotopic by a differentiable homotopy.

**Proof.** See [Hirs76] for a precise statement and proofs.

Proving the results we just mentioned about differentiable structures is beyond the scope of this book. See Notes 1–4 in Section 6.5 for a few other related comments. It follows from Theorem 4.4.5 that locally the imbedding of a k-dimensional smooth manifold $\mathbf{M}^k$ in $\mathbf{R}^n$ looks like the imbedding of $\mathbf{R}^k$ in $\mathbf{R}^n$. The next two theorems relate the manifolds defined in Section 8.3 to the abstract manifolds we are studying now. Combined, the theorems say that there is no difference between the two types of spaces. It is simply a case of two different ways of looking at the same thing.

**8.8.6. Theorem.** A subset of $\mathbf{R}^n$ is a differentiable manifold in the sense of Section 8.3 if and only if it is a differentiable submanifold of $\mathbf{R}^n$ in the sense of this section.

**Proof.** This is a consequence of the Inverse Function Theorem.

**8.8.7. Theorem.** Every abstract manifold $\mathbf{M}^k$ can be imbedded in some Euclidean space $\mathbf{R}^n$.

**Proof.** See [Miln58], [Munk61], or [Hirs76].

The proof of Theorem 8.8.7 is not very hard if one is happy with any n. A classical result of H. Whitney, proved in 1936, states that $n = 2k + 1$ suffices. In fact, one can improve this to $n = 2k$, but this is even more difficult.

Among other things, Theorems 8.8.6 and 8.8.7 imply that Theorem 8.3.3 also holds for abstract manifolds. The reader may also well ask why one bothers to give the more complicated intrinsic definition of a manifold and not simply stick to submanifolds of $\mathbf{R}^n$. To repeat, the abstraction enables one to see essential aspects more clearly. Manifolds have lots of intrinsic properties that do not depend on any particular imbedding. Nevertheless, the usual way that one proceeds with problems dealing with abstract manifolds $\mathbf{M}^k$ is to divide those problem into lots of local problems that can then be translated into problems in $\mathbf{R}^k$ using a coordinate neighborhood. In this way, an understanding of abstract manifolds largely reduces to a good understanding of Euclidean space (in the same way that an understanding of vector-valued functions largely reduces to an understanding of real-valued functions).

The last topic in this section is to show that abstract manifolds also have intrinsically defined tangent spaces associated to every point. There are two basic approaches to defining tangent vectors for a manifold $\mathbf{M}^k$. Let $\mathbf{p} \in \mathbf{M}^k$.

**The equivalence class of vectors approach to tangent vectors at p**: This approach is based on what we did in Section 8.3. If we think of $\mathbf{R}^k$ as a submanifold of $\mathbf{R}^k$, then the set of tangent vectors of curves in $\mathbf{R}^k$ through the point $\mathbf{p}$ is just $\mathbf{R}^k$. In other words, the space of all tangent vectors at all points looks like $\mathbf{R}^k \times \mathbf{R}^k$ with the tangent vectors at $\mathbf{p}$ being the set $\mathbf{p} \times \mathbf{R}^k$. For an abstract manifold we shall do something similar for each coordinate neighborhood of $\mathbf{p}$. For every coordinate neighborhood $(\mathbf{U},\varphi)$ for $\mathbf{p}$ and $\mathbf{a} \in \mathbf{R}^k$, we would like to call $(\mathbf{U},\varphi,\mathbf{a})$ a tangent vector at $\mathbf{p}$. The only problem is that $\mathbf{p}$ will in general belong to many different coordinate neighborhoods. If $(\mathbf{V},\psi)$ is another coordinate neighborhood of $\mathbf{p}$ and $\mathbf{b} \in \mathbf{R}^k$, then we would also have called $(\mathbf{V},\psi,\mathbf{b})$ a tangent vector at $\mathbf{p}$. We need an equivalence relation for these tuples. Consider the diagram

$$\begin{tikzcd}
 & \mathbf{U}\cap\mathbf{V} \subset \mathbf{M}^k \arrow[dl, "\varphi"'] \arrow[dr, "\psi"] & \\
\varphi(\mathbf{U}\cap\mathbf{V}) \subset \mathbf{R}^k & & \psi(\mathbf{U}\cap\mathbf{V}) \subset \mathbf{R}^k \\
(u_1,u_2,\ldots,u_k) \arrow[rr, "\psi\circ\varphi^{-1}"] & & (v_1,v_2,\ldots,v_k).
\end{tikzcd}$$

Define a relation ~ as follows:

$$(\mathbf{U},\varphi,\mathbf{a}) \sim (\mathbf{V},\psi,\mathbf{b})$$

if

$$\mathbf{b} = D\big(\psi \circ \varphi^{-1}\big)(\varphi(\mathbf{p}))(\mathbf{a}). \tag{8.12}$$

It is easy to check that ~ is an equivalence relation. Furthermore, if

$$\mathbf{a} = (a_1, a_2, \ldots, a_k) \quad \text{and} \quad \mathbf{b} = (b_1, b_2, \ldots, b_k)$$

and if one thinks of each $v_i$ as a differentiable function of $u_1$, $u_2$, . . . , and $u_k$, then equation (8.12) is equivalent to the set of equations

$$b_i = \sum_{j=1}^{k} a_j \frac{\partial v_i}{\partial u_j}. \tag{8.13}$$

**Definition.** An equivalence class of tuples $(\mathbf{U},\varphi,\mathbf{a})$ with respect to the relation ~, denoted by $[\mathbf{U},\varphi,\mathbf{a}]$ or $[\mathbf{U},\varphi,\mathbf{a}]_{\mathbf{p}}$, is called a *tangent vector* of $\mathbf{M}^k$ at $\mathbf{p}$. One calls $a_i$ the *ith component* of the tangent vector $[\mathbf{U},\varphi,\mathbf{a}]$ in the coordinate neighborhood $(\mathbf{U},\varphi)$. The *tangent space* of $\mathbf{M}^k$ at $\mathbf{p}$, $T_{\mathbf{p}}(\mathbf{M}^k)$, is the set of all tangent vectors to $\mathbf{M}^k$ at $\mathbf{p}$. The set $T_{\mathbf{p}}(\mathbf{M}^k)$ is made into a vector space by defining

$$[\mathbf{U},\varphi,\mathbf{a}] + [\mathbf{U},\varphi,b] = [\mathbf{U},\varphi,\mathbf{a}+\mathbf{b}]$$
$$c[\mathbf{U},\varphi,\mathbf{a}] = [\mathbf{U},\varphi,c\mathbf{a}].$$

One can show that $T_{\mathbf{p}}(\mathbf{M}^k)$ is in fact a k-dimensional vector space. Furthermore, we can tie our new definition to the earlier one where we defined tangent vectors in terms of curves.

**Definition.** If $\gamma:(c-\varepsilon,c+\varepsilon) \to \mathbf{U}$ (or $\gamma:[c,c+\varepsilon) \to \mathbf{U}$ if $\mathbf{p} \in \partial\mathbf{M}$) is a curve with $\gamma(c) = \mathbf{p}$, then

$$[\mathbf{U},\varphi,\mathbf{D}(\varphi\circ\gamma)(c)(1)] = \left[\mathbf{U},\varphi,\left(\frac{d}{dt}(u_1\circ\gamma)(c),\ldots,\frac{d}{dt}(u_k\circ\gamma)(c)\right)\right] \tag{8.14}$$

is called the *tangent vector* of $\gamma(t)$ at c.

Tangent vectors at points of curves in a manifold are well-defined tangent vectors of the manifold that depend only on the curve near the points and not on the choice of coordinate neighborhood $(\mathbf{U},\varphi)$. They can be used to define standard bases for the tangent spaces. Define curves

$$\gamma_i:\ (-\varepsilon,\varepsilon) \to \mathbf{U}$$

by

$$\gamma_i(t) = \varphi^{-1}(\varphi(\mathbf{p}) + t\mathbf{e}_i).$$

**Notation.** Denote the tangent vector to $\gamma_i(t)$ at 0 by $\mathbf{e}_{i,\mathbf{U}}$.

It is easy to show that the vectors $\mathbf{e}_{1,\mathbf{U}}, \mathbf{e}_{2,\mathbf{U}}, \ldots, \mathbf{e}_{k,\mathbf{U}}$ form a basis for $T_{\mathbf{p}}(\mathbf{M}^k)$ (Exercise 8.8.2). They are the natural basis of $T_{\mathbf{p}}(\mathbf{M}^k)$ with respect to the coordinate neighborhood $(\mathbf{U},\varphi)$ and the current definition of tangent vectors.

**Definition.** Let $\mathbf{M}^k$ and $\mathbf{W}^m$ be differentiable manifolds and let $f:\mathbf{M}^k \to \mathbf{W}^m$ be a differentiable map. If $\mathbf{p} \in \mathbf{M}^k$, then define a map

$$Df(\mathbf{p}):\ T_{\mathbf{p}}(\mathbf{M}^k) \to T_{f(\mathbf{p})}(\mathbf{W}^m) \tag{8.15a}$$

by

$$Df(\mathbf{p})([\mathbf{U},\varphi,\mathbf{a}]) = [\mathbf{V},\psi,\mathbf{b}], \tag{8.15b}$$

where $(\mathbf{U},\varphi)$ and $(\mathbf{V},\psi)$ are coordinate neighborhoods of $\mathbf{p}$ and $f(\mathbf{p})$, respectively and

$$\mathbf{b} = D\left(\psi \circ f \circ \varphi^{-1}\right)(\varphi(\mathbf{p}))(\mathbf{a}). \tag{8.15c}$$

The map $Df(\mathbf{p})$ is called the *derivative* of f at $\mathbf{p}$.

One can show that $Df(\mathbf{p})$ is a well-defined linear transformation.

In classical terminology, tangent vectors are called contravariant tensors. This was motivated by the definition that we just gave. The terms "covariant" and "contravariant" have to do with how quantities transform under a change of coordinates. See [Spiv70a] for a discussion of why this classical terminology is unfortunate and why tangent vectors should really be called covariant tensors. The fact is that the terminology has been around for so long, so that no one has dared to change it.

**The linear functional approach to tangent vectors at p:** First, define

$$F(\mathbf{p}) = \{f : \mathbf{U} \to \mathbf{R} \mid \mathbf{U} \text{ is an open neighborhood of } \mathbf{p} \text{ in } \mathbf{M} \text{ and f is differentiable}\}.$$

**Definition.** A *tangent vector* of $\mathbf{M}^k$ at $\mathbf{p}$ is a map

$$X\colon\ F(\mathbf{p}) \to \mathbf{R}$$

satisfying

(1) $X(af + bg) = aX(f) + bX(g)$ for all $f, g \in F(\mathbf{p})$ and $a, b \in \mathbf{R}$.
(2) $X(fg) = X(f)g(\mathbf{p}) + f(\mathbf{p})X(g)$ for all $f, g \in F(\mathbf{p})$.

(The domain of $af + bg$ and $fg$ is the intersection of the domain of f and the domain of g.) The *tangent space* of $\mathbf{M}^k$ at $\mathbf{p}$, $T_{\mathbf{p}}(\mathbf{M}^k)$, is the set of all tangent vectors to $\mathbf{M}^k$ at $\mathbf{p}$.

**Note.** A map X satisfying properties (1) and (2) is usually called a *derivation* because it acts like a derivative.

This definition is motivated by the following observations. Let us return to Section 8.3 and assume that manifolds are subsets of Euclidean space. If $\gamma\colon[a,b] \to \mathbf{M}^k$ is a curve with $\gamma(c) = \mathbf{p}$, then $\gamma$ induces a map

$$\gamma_*\colon\ F(\mathbf{p}) \to \mathbf{R}$$

defined by

$$\gamma_*(f) = \frac{d}{dt}(f \circ \gamma)(c).$$

The map $\gamma_*$ is just the directional derivative of f at $\mathbf{p}$ in the direction $\gamma'(c)$. One can show that $\gamma_*$ depends only on the derivative of $\gamma(t)$ at c. In other words, one can think

of a tangent vector as something that specifies a direction in which one can take a derivative. Here are some properties of these tangent vectors:

(1) Let $c_{\mathbf{U}}$ denote the constant map on the neighborhood $\mathbf{U}$ with value c. Then

$$X(c_{\mathbf{U}}) = 0.$$

This follows from the identities

$$X(c_{\mathbf{U}}) = cX(1_{\mathbf{U}}) = cX(1_{\mathbf{U}} \cdot 1_{\mathbf{U}}) = c(X(1_{\mathbf{U}}) + X(1_{\mathbf{U}})) = 2cX(1_{\mathbf{U}}).$$

(2) Let $f: \mathbf{U} \to \mathbf{R}$ and $\mathbf{V} \subseteq \mathbf{U}$. Then

$$X(f|\mathbf{V}) = X(f \cdot 1_{\mathbf{V}}) = X(f) + f(\mathbf{p})X(1_{\mathbf{V}}) = X(f),$$

that is, X(f) depends only on the local behavior of f and not on its domain.

(3) Let $(\mathbf{U},\varphi)$ be a coordinate neighborhood of $\mathbf{p}$ and let $\varphi(\mathbf{q}) = (u_1(\mathbf{q}),u_2(\mathbf{q}), \ldots ,u_k(\mathbf{q}))$, $\mathbf{q} \in \mathbf{U}$. If

$$f \in F(\mathbf{p}),$$

then define

$$\frac{\partial}{\partial u_i}(f) = \frac{\partial}{\partial x_i}(f \circ \varphi^{-1})(\varphi(\mathbf{p})). \tag{8.16}$$

It is easy to see that $\partial/\partial u_i$ is a tangent vector at $\mathbf{p}$ called the *partial derivative with respect to $u_i$*. Furthermore, $\partial/\partial u_1, \partial/\partial u_2, \ldots, \partial/\partial u_k$ are a basis for the tangent space $T_{\mathbf{p}}(\mathbf{M}^k)$ (Exercise 8.8.3). (In contrast to other results that hold for $C^r$ manifolds also, the basis property of the $\partial/\partial u_i$ needs $C^\infty$, because to show that the $\partial/\partial u_i$ span, one needs Lemma 4.6.2. See [BisC64].) It follows that every tangent vector X can be written uniquely in the form

$$X = \sum_{j=1}^{k} a_j \frac{\partial}{\partial u_j}.$$

One calls $a_i$ the *ith component* of the tangent vector X.

(4) Let $(\mathbf{V},\psi)$ be another coordinate neighborhood of $\mathbf{p}$ and let $\psi(\mathbf{q}) = (v_1(\mathbf{q}),v_2(\mathbf{q}), \ldots ,v_k(\mathbf{q}))$, $\mathbf{q} \in \mathbf{V}$. If we express X with respect to the basis $\partial/\partial v_j$, that is, if

$$X = \sum_{j=1}^{k} b_j \frac{\partial}{\partial v_j},$$

then

$$a_i = X(u_i) = \sum_{j=1}^{k} b_j \frac{\partial u_i}{\partial v_j}. \tag{8.17}$$

We see that this matches equation (8.13) and the two definitions of tangent vectors really amount to the same thing. In fact, the correspondence

$$\sum_{j=1}^{k} a_j \frac{\partial}{\partial u_j} \leftrightarrow \sum_{j=1}^{k} a_j \mathbf{e}_{j,\mathbf{U}} \tag{8.18}$$

defines the natural isomorphism between the two vector spaces that are called the tangent space to the manifold.

**Definition.** Let $\mathbf{M}^k$ and $\mathbf{W}^m$ be differentiable manifolds and let $f: \mathbf{M}^k \to \mathbf{W}^m$ be a differentiable map. If $\mathbf{p} \in \mathbf{M}^k$, then define a map

$$Df(\mathbf{p}):\ T_p\left(\mathbf{M}^k\right) \to T_{f(\mathbf{p})}\left(\mathbf{W}^m\right) \tag{8.19a}$$

by

$$(Df(\mathbf{p})(X))(g) = X(g \circ f) \tag{8.19b}$$

for every $X \in T_{\mathbf{p}}(\mathbf{M}^k)$ and $g \in F(f(\mathbf{p}))$. The map $Df(\mathbf{p})$ is called the *derivative* of f at $\mathbf{p}$.

Just like with the previous equivalence class of vectors definition, one can show that $Df(\mathbf{p})$ is a well-defined linear transformation.

No matter which definition of tangent vectors one uses, given a map f between differentiable manifolds, the *rank of f at a point* $\boldsymbol{p}$ is the rank of its derivative $Df(\mathbf{p})$.

Let us summarize the main points that we covered in this section. We defined abstract manifolds, tangent vectors, when a map is differentiable, and the derivative of a map. One can show that for submanifolds of Euclidean space the notions of tangent vectors, the derivative of a map, and the rank of a map are compatible with those given in Section 8.4.

## 8.9 Vector Bundles

Bundles over a space were introduced in Section 7.4.2. The basic concept consisted of three pieces, a total space, a projection map onto a base space, and a local triviality condition (each point of the base space had a neighborhood over which the bundle looked like a product of the base neighborhood and another space called the fiber). In Section 7.4.2 we concentrated on a very special case, that of covering spaces, where the fiber was a discrete space. In this section we look at the case of where the fiber is a vector space. Even though covering spaces are really part of the same general topic of "fiber" bundles, for historical reasons the notation differs slightly between the two. We shall now switch to the notation used for vector bundles. (In Section 7.4.2 we used the expression "bundle **over a space**" to emphasize that the base space was not part of the definition of "bundle", which it will be here.)

**Definition.** An *n-dimensional (real) vector bundle*, or *n-plane bundle*, or simply *vector bundle* if the dimension is unimportant, is a triple $\xi = (\mathbf{E}(\xi),\pi_\xi,\mathbf{B}(\xi)) = (\mathbf{E},\pi,\mathbf{B})$ satisfying

(1) $\mathbf{E}$ and $\mathbf{B}$ are topological Hausdorff spaces.
(2) The map $\pi:\mathbf{E} \to \mathbf{B}$ is continuous and onto.
(3) For each $\mathbf{b} \in \mathbf{B}$, $\pi^{-1}(\mathbf{b})$ has the structure of an n-dimensional (real) vector space.
(4) (Local triviality) For each $\mathbf{b} \in \mathbf{B}$, there is an open neighborhood $\mathbf{U_b}$ of $\mathbf{b}$ and a homeomorphism

$$\varphi_\mathbf{b}:\ \mathbf{U_b} \times \mathbf{R}^n \to \pi^{-1}(\mathbf{U_b})$$

such that

$$\varphi_b\,|\,\mathbf{b}' \times \mathbf{R}^n : \mathbf{b}' \times \mathbf{R}^n \to \pi^{-1}(\mathbf{b}')$$

is a vector space isomorphism for all $\mathbf{b}' \in \mathbf{U_b}$.

$\mathbf{E}$ is called the *total space*, $\pi$ is called the *projection*, and $\mathbf{B}$ is called the *base space* for $\xi$. The space $\pi^{-1}(\mathbf{b})$ is called the *fiber* of $\xi$ over $\mathbf{b}$. The pair $(\varphi_\mathbf{b},\mathbf{U_b})$ is called a *local coordinate chart* for the vector bundle. A one-dimensional vector bundle is often called a *line bundle*. One sometimes refers to $\xi$ as a *vector bundle over* $\boldsymbol{B}$.

**8.9.1. Example.** If $\mathbf{B}$ is a topological space and $\pi:\mathbf{B} \times \mathbf{R}^n \to \mathbf{B}$ is the projection onto the first factor, then $\xi = (\mathbf{B} \times \mathbf{R}^n,\pi,\mathbf{B})$ is clearly an n-plane bundle called the *product n-plane bundle over* $\boldsymbol{B}$.

Example 8.9.1 shows that there are lots of vector bundles, but the theory would not be very interesting if they all were just product bundles. We shall see examples of other bundles shortly, but we need a few more definitions first.

**Definition.** Let $\xi = (\mathbf{E},\pi,\mathbf{B})$ be an n-plane bundle and let $\mathbf{A} \subseteq \mathbf{B}$. The *restriction of* $\xi$ *to* $\boldsymbol{A}$, $\xi|\mathbf{A}$, is the n-plane bundle $\xi|\mathbf{A} = (\pi^{-1}(\mathbf{A}),\pi|\pi^{-1}(\mathbf{A}),\mathbf{A})$.

Showing that $\xi|\mathbf{A}$ is an n-plane bundle is an easy exercise.

**Definition.** Let $\xi = (\mathbf{E},\pi,\mathbf{B})$ be an n-plane bundle. A *cross-section* of $\xi$ is a continuous map

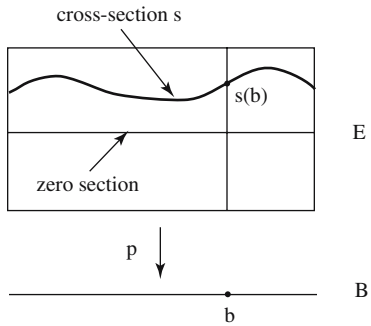
$$\mathbf{s}:\ \mathbf{B} \to \mathbf{E}$$

so that

$$\pi \circ s = 1_\mathbf{B}\,,$$

that is, $s(\mathbf{b})$ belongs to the fiber $\pi^{-1}(\mathbf{b})$ for every $\mathbf{b} \in \mathbf{B}$. The *zero cross-section* is the cross-section s where $s(\mathbf{b})$ is the zero vector in $\pi^{-1}(\mathbf{b})$ for every $\mathbf{b} \in \mathbf{B}$. A cross-section s is said to be *nonzero* if $s(\mathbf{b})$ is a nonzero vector in the vector space $\pi^{-1}(\mathbf{b})$ for every

**Figure 8.28.** Cross-sections.




$\mathbf{b} \in \mathbf{B}$. The *support* of a cross-section s, denoted by support(s), is defined to be the set of $\mathbf{b} \in \mathbf{B}$ with $s(\mathbf{b}) \neq \mathbf{0}$. Two cross-sections $s_1$ and $s_2$ are said to be *linearly independent* if $s_1(\mathbf{b})$ and $s_2(\mathbf{b})$ are linearly independent vectors in $\pi^{-1}(\mathbf{b})$ for every $\mathbf{b} \in \mathbf{B}$.

See Figure 8.28. It is easy to see that the set of cross-sections of a vector bundle $\xi$ is actually a vector space. In fact, if s is a cross-section of $\xi$ and f is a real-valued function on the base space $\mathbf{B}$ of $\xi$, then we can define a new cross-section fs for $\xi$ by means of the obvious formula

$$(fs)(\mathbf{b}) = f(\mathbf{b})s(\mathbf{b}), \quad \text{for all} \quad \mathbf{b} \in \mathbf{B}.$$

**Note.** Every vector bundle has the zero cross-section. One often identifies the base space of a vector bundle with the image of this zero cross-section in the total space, namely, the space of zero vectors in all the fibers.

Next, we define what is meant by a map between vector bundles. Such a map should preserve fibers and the vector space structure.

**Definition.** Let $\xi_i = (\mathbf{E}_i, \pi_i, \mathbf{B}_i)$ be vector bundles (of possibly different dimensions). A *vector bundle map* $F: \xi_1 \to \xi_2$ is a pair of maps $(\tilde{f}, f)$, so that

(1) the diagram

$$\begin{array}{ccc} \mathbf{E}_1 & \xrightarrow{\tilde{f}} & \mathbf{E}_2 \\ \pi_1 \downarrow & & \downarrow \pi_2 \\ \mathbf{B}_1 & \xrightarrow{f} & \mathbf{B}_2 \end{array}$$

commutes, that is, $\pi_2 \circ \tilde{f} = f \circ \pi_1$, and

(2) the *fiber maps*

$$\tilde{f}_F(\mathbf{b}_1) = \tilde{f}\big|_{\pi_1^{-1}(\mathbf{b}_1)}: \ \pi_1^{-1}(\mathbf{b}_1) \to \pi_2^{-1}(f(\mathbf{b}_1))$$

are linear transformations with respect to the vector space structure on each fiber for all $\mathbf{b}_1 \in \mathbf{B}_1$.

The map F is called a *vector bundle isomorphism* if f is a homeomorphism and $\tilde{f}$ is a vector space isomorphism on each fiber (or, equivalently, $\tilde{f}$ is a homeomorphism between $\mathbf{E}_1$ and $\mathbf{E}_2$). We say that $\xi_1$ and $\xi_2$ are *isomorphic* vector bundles, and write $\xi_1 \approx \xi_2$, if there exists a vector bundle isomorphism between them. An n-plane bundle $\xi = (\mathbf{E},\pi,\mathbf{B})$ that is isomorphic to the product bundle ($\mathbf{B} \times \mathbf{R}^n$, projection onto $\mathbf{B}$,$\mathbf{B}$) is called a *trivial vector bundle*.

Clearly, if $F = (\tilde{f},f)$ is a vector bundle isomorphism, then $F^{-1} = (\tilde{f}^{-1},f^{-1})$ is a vector bundle isomorphism called the *inverse* of F. One can also compose vector bundle maps.

Now, one way of thinking of vector bundles is as locally trivial bundles with fiber a vector space, that is, if $\xi$ is an n-plane bundle over a space $\mathbf{B}$, then there will exist a covering of $\mathbf{B}$ by open sets $\mathbf{U}$, so that $\xi|\mathbf{U}$ is trivial. To put is yet another way, an n-plane bundle over a space $\mathbf{B}$ consists of a collection of product bundles ($\mathbf{U} \times \mathbf{R}^n$,projection onto $\mathbf{U}$,$\mathbf{U}$) for open sets $\mathbf{U}$ in $\mathbf{B}$ that are glued together along their fibers using maps in GL(n,$\mathbf{R}$). We shall make this clearer later when we discuss the tangent bundle of a manifold. The generalization of this way of looking at a vector bundle is what is called a *fiber bundle* (also written *"fibre" bundle*) where we allow an arbitrary space to be the fiber except that one is also explicitly given a group which acts on the fibers. In our case, this group would be GL(n,$\mathbf{R}$).

Returning to the topic of cross-sections, an interesting question is whether or not a vector bundle has a nonzero cross-section. Trivial bundles (other than the 0-dimensional ones) certainly have lots of nonzero cross-sections. In fact, a trivial n-plane bundles has n linearly independent cross-sections (Exercise 8.9.1). Therefore, a vector bundle that has no nonzero cross-section cannot be trivial and this becomes one of the tests for triviality.

It is time to give an example of a nontrivial vector bundle.

**8.9.2. Example.** Our nontrivial vector bundle is easy to describe in rough terms, although it will take a little more work to explain rigorously. Basically, we are talking about an *"open" Moebius strip* thought of as the total space of a line bundle ($\mathbf{E}$,p,$\mathbf{S}^1$) over the circle. See Figure 8.29. In a sense, we are simply removing the boundary of

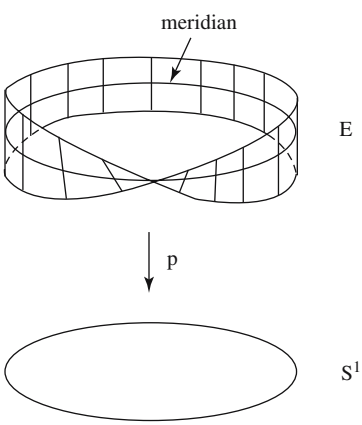



**Figure 8.29.** The open Moebius strip line bundle.

the Moebius strip defined in Chapter 6, but we shall modify our earlier construction to match the vector bundle structure better. To this end, let the total space **E** be the space $[0,\pi] \times \mathbf{R}$ with the two ends glued together after giving one end a 180° twist, that is, the point (0,t) is identified with the point (π,–t). The base space is the center line, or *meridian*, $[0,\pi] \times 0$ with the end point (0,0) glued to the end point (π,0). What we get is a circle that is identified with $\mathbf{S}^1$ and we have the obvious projection map p which maps each fiber $c \times \mathbf{R}$ to c. The space **E** is obviously not homeomorphic to $\mathbf{S}^1 \times \mathbf{R}$ and so the bundle is not trivial. This is the quick and dirty description of the line bundle we are after, but filling in the missing details would be a little messy. Therefore, we shall now describe a quite different construction for the "same" bundle, one that leads to a nice generalization in Section 8.13.

Let

$$p\colon\ \mathbf{S}^1 \to \mathbf{P}^1$$

be the standard 2-fold covering of $\mathbf{P}^1$, which maps every point $\mathbf{q} \in \mathbf{S}^1$ into the equivalence class

$$[\mathbf{q}] = \{\mathbf{q}, -\mathbf{q}\} \in \mathbf{P}^1$$

**Definition.** The *canonical line bundle* $\gamma = (\mathbf{E},\pi,\mathbf{P}^1)$ over $\mathbf{P}^1$ is defined as follows:

(1) $\mathbf{E} = \{([\mathbf{q}],t\mathbf{q}) \in \mathbf{P}^1 \times \mathbf{R}^2 \mid t \in \mathbf{R}\}$.
(2) $\pi([\mathbf{q}],t\mathbf{q}) = [\mathbf{q}]$.

To show that γ is a vector bundle, we must show that the fibers have a vector space structure and that the bundle is locally trivial. Since the fibers of this bundle are just the lines through the origin in $\mathbf{R}^2$, we can obviously consider them as one-dimensional vector spaces. To prove the locally triviality property define sets

$$\tilde{\mathbf{U}}_i = \{\mathbf{q} = (q_1, q_2) \in \mathbf{S}^1 \subset \mathbf{R}^2 \mid q_i > 0\} \subset \mathbf{S}^1 \quad \text{and} \quad \mathbf{U}_i = p(\tilde{\mathbf{U}}_i) \subset \mathbf{P}^1. \tag{8.26}$$

The sets $\mathbf{U}_i$ are open sets whose union is $\mathbf{P}^1$. Define homeomorphisms

$$\varphi_i\colon\ \mathbf{U}_i \times \mathbf{R} \to \pi^{-1}(\mathbf{U}_i)$$

by

$$\varphi_i(\mathbf{x}, t) = (\mathbf{x}, t\mathbf{q}),$$

where $\mathbf{x} = [\mathbf{q}]$ and the representative $\mathbf{q}$ for $\mathbf{x}$ is chosen so that $q_i > 0$. It is easy to check that the maps $\varphi_i$ are well-defined homeomorphisms because the sets $\tilde{\mathbf{U}}_i$ do not contain antipodal points. This finishes the proof that γ is a line bundle. Exercise 8.9.2 asks the reader to show that γ is isomorphic to the open Moebious strip bundle described above.

Finally, we prove that the line bundle γ is not trivial by showing that it does not admit any nonzero cross-section. Suppose that γ had a nonzero cross-section s. It would follow that the map

$$s \circ p\colon \mathbf{S}^1 \to \mathbf{E}$$

would have the form

$$(s \circ p)(\mathbf{q}) = ([\mathbf{q}], \alpha(\mathbf{q})\mathbf{q}), \tag{8.27}$$

for some continuous map

$$\alpha\colon \mathbf{S}^1 \to \mathbf{R} \text{ with } \alpha(-\mathbf{q}) = -\alpha(\mathbf{q}).$$

(Exercise 8.9.3). Such a map $\alpha$ takes on both positive and negative values. Since $\mathbf{S}^1$ is connected, the intermediate value theorem implies that $\alpha$ must be zero somewhere, which contradicts the hypothesis that s was a nonzero cross-section. This finishes Example 8.9.2.

Here are some important constructions defined for vector bundles.

**Definition.** Let $\xi_i = (\mathbf{E}_i, \pi_i, \mathbf{B}_i)$ be vector bundles. The vector bundle $\xi_1 \times \xi_2 = (\mathbf{E}_1 \times \mathbf{E}_2, \pi_1 \times \pi_2, \mathbf{B}_1 \times \mathbf{B}_2)$ is called the *vector bundle product of* $\xi_1$ *and* $\xi_2$.

It is trivial to show that $\xi_1 \times \xi_2$ is in fact a vector bundle.

**Definition.** Let $\varphi = (\mathbf{E}, \pi, \mathbf{B})$ be an n-plane bundle and let $f\colon \mathbf{B}_1 \to \mathbf{B}$ be a map. Define an n-plane bundle $f^*\xi = (\mathbf{E}_1, \pi_1, \mathbf{B}_1)$ over $\mathbf{B}_1$ as follows:

(1) $\mathbf{E}_1 = \{(\mathbf{b}_1, \mathbf{e}) \in \mathbf{B}_1 \times \mathbf{E} \mid f(\mathbf{b}_1) = \pi(\mathbf{e})\}$
(2) $\pi_1(\mathbf{b}_1, \mathbf{e}) = \mathbf{b}_1$
(3) The vector space structure for each fiber $\pi_1^{-1}(\mathbf{b}_1)$ is defined by

$$r(\mathbf{b}_1, \mathbf{e}) + s(\mathbf{b}_1, \mathbf{e}') = (\mathbf{b}_1, r\mathbf{e} + s\mathbf{e}'), \text{ for } r, s \in \mathbf{R}.$$

(4) Let $\mathbf{b}_1 \in \mathbf{B}_1$ and let

$$\varphi\colon \mathbf{U} \times \mathbf{R}^n \to \pi^{-1}(\mathbf{U})$$

be a local coordinate chart for $\xi$ over a neighborhood $\mathbf{U}$ of $f(\mathbf{b}_1)$. If $\mathbf{V} = f^{-1}(\mathbf{U})$, then the map

$$\varphi_1\colon U \times \mathbf{R}^n \to \pi_1^{-1}(\mathbf{V}),$$

defined by

$$\varphi_1(\mathbf{b}_1', \mathbf{v}) = (\mathbf{b}_1', \varphi(f(\mathbf{b}_1'), \mathbf{v})).$$

is a local coordinate chart for $\xi_1$ over $\mathbf{V}$.

The vector bundle $f^*\xi$ is called the *induced (vector) bundle* over $\mathbf{B}_1$, the *(vector) bundle* over $\mathbf{B}_1$ *induced by f*, or the *pullback (vector) bundle*. Define

$$\tilde{f}\colon\ \mathbf{E}_1 \to \mathbf{E}$$

by

$$\tilde{f}(\mathbf{b}_1, \mathbf{e}) = \mathbf{e}.$$

The bundle map $(\tilde{f},f): f^*\xi \to \xi$ is called the *canonical (vector bundle) map* from $f^*\varphi$ to $\xi$.

**8.9.3. Lemma.** $f^*\xi$ is a well-defined n-plane bundle and $(\tilde{f},f)$ is a vector bundle map.

**Proof.** Easy.

We collect the main facts about induced bundles in the next theorem.

**8.9.4. Theorem.**

(1) If $(\tilde{f},f): \eta \to \xi$ is a vector bundle map between two n-plane bundles $\eta$ and $\xi$, then the induced vector bundle $f^*\xi$ is isomorphic to $\eta$.
(2) If $\varphi = (\mathbf{E},\pi,\mathbf{B})$ is a trivial vector bundle over $\mathbf{B}$ and if $f: \mathbf{B}_1 \to \mathbf{B}$ is a map, then $f^*\varphi$ is a trivial vector bundle over $\mathbf{B}_1$.
(3) Let $\xi = (\mathbf{E},\pi,\mathbf{B})$ be a vector bundle. If $\mathbf{B}_1$ is a paracompact space and if f, g: $\mathbf{B}_1 \to \mathbf{B}$ are homotopic maps, then the induced bundles $f^*\xi$ and $g^*\xi$ are isomorphic.

**Proof.** To prove (1) show that the vector bundle map

$$(\tilde{g}, 1_{\mathbf{B}(\eta)})\colon\ \eta \to f^*\xi$$

defined by

$$\tilde{g}(\mathbf{e}) = (\pi_\eta(\mathbf{e}), \tilde{f}(\mathbf{e}))$$

is the desired isomorphism. Fact (2) is easy. For fact (3) see [Huse66].

**8.9.5. Corollary.** Every vector bundle over a contractible paracompact space $\mathbf{B}$ is trivial.

**Proof.** Let f be the identity map on $\mathbf{B}$ and let $g: \mathbf{B} \to \mathbf{B}$ be a constant map. If $\xi$ is any vector bundle over $\mathbf{B}$, then $f^*\xi$ is easily seen to be isomorphic to $\xi$ and $g^*\xi$ is trivial by Theorem 8.9.4(2). Since f is homotopic to g, the Corollary now follows from Theorem 8.9.4(3).

Next, we show how vector bundles over a space can be added.

**Definition.** Let $\xi_1 = (\mathbf{E}_1,\pi_1,\mathbf{B})$ be an n-plane bundle and $\xi_2 = (\mathbf{E}_2,\pi_2,\mathbf{B})$ an m-plane bundle over the same base space $\mathbf{B}$. Define the *Whitney sum* of $\xi_1$ and $\xi_2$, denoted by $\xi_1 \oplus \xi_2$, to be the vector bundle $(\mathbf{E},\pi,\mathbf{B})$, where

(1) $\mathbf{E} = \{(\mathbf{e}_1,\mathbf{e}_2) \in \mathbf{E}_1 \times \mathbf{E}_2 \mid \pi_1(\mathbf{e}_1) = \pi_2(\mathbf{e}_2)\}$,
(2) $\pi(\mathbf{e}_1,\mathbf{e}_2) = \pi_1(\mathbf{e}_1) = \pi_2(\mathbf{e}_2)$,
(3) the vector space structure for each fiber $\pi^{-1}(\mathbf{b}) = \pi_1^{-1}(\mathbf{b}) \times \pi_2^{-1}(\mathbf{b})$ is just the direct sum vector space structure, and
(4) if $\mathbf{b} \in \mathbf{B}$ and if

$$\varphi_1\colon \mathbf{U} \times \mathbf{R}^n \to \pi_1^{-1}(\mathbf{U}) \quad \text{and} \quad \varphi_2\colon \mathbf{U} \times \mathbf{R}^m \to \pi_2^{-1}(\mathbf{U})$$

are local coordinate charts for $\xi_1$ and $\xi_2$, respectively, over a neighborhood $\mathbf{U}$ of $\mathbf{b}$, then

$$\varphi\colon \mathbf{U} \times \mathbf{R}^{n+m} \to \pi^{-1}(\mathbf{U}),$$

defined by

$$\varphi(\mathbf{b}',(\mathbf{v},\mathbf{w})) = (\varphi_1(\mathbf{b}',\mathbf{v}),\varphi_2(\mathbf{b}',\mathbf{w})),\ \mathbf{v} \in \mathbf{R}^n, \mathbf{w} \in \mathbf{R}^m,$$

is a local coordinate chart for $\xi_1 \oplus \xi_2$ over $\mathbf{U}$.

**8.9.6. Lemma.** $\xi_1 \oplus \xi_2$ is a well-defined (n + m)-dimensional vector bundle.

**Proof.** Easy.

One very useful notion for vector spaces is that of an inner product because then one can talk about the length of vectors and whether two are orthogonal. It is convenient to have these concepts for vector bundles.

**Definition.** A *Riemannian metric* on a vector bundle $\xi$ is a continuous function

$$\mu\colon \mathbf{E}(\xi) \to \mathbf{R}$$

such that the restriction of $\mu$ to each fiber is a positive definite quadratic form.

**8.9.7. Theorem.** Every vector bundle over a paracompact space admits a Riemannian metric.

**Proof.** See [Spiv70a].

**Note.** Because of the equivalence between quadratic forms and symmetric bilinear maps, a Riemannian metric on a vector bundle $\xi$ is sometimes defined to be a continuous function

$$<,> : \mathbf{E}(\xi) \oplus \mathbf{E}(\xi) \to \mathbf{R}$$

with the property that <,> is a positive definite symmetric bilinear map (inner product) on each fiber. We shall feel free to switch between these two equivalent definitions of a Riemannian metric.

Once a vector bundle has a Riemannian metric, then the terms "length", "orthogonality", "angle", etc., will all make sense with regard to vectors in a fiber. One can also define two natural sub-bundles.

**Definition.** Let $\xi = (\mathbf{E},\pi,\mathbf{X})$ be a vector bundle with Riemannian metric $\mu$. Define

$$\mathbf{E_D} = \{\mathbf{e} \in \mathbf{E} \mid \mu(\mathbf{e}) \le 1\} \quad \text{and} \quad \mathbf{E_S} = \{\mathbf{e} \in \mathbf{E} \mid \mu(\mathbf{e}) = 1\}.$$

The bundles $\mathbf{D}(\xi) = (\mathbf{E_D},\pi|\mathbf{E_D},\mathbf{X})$ and $\mathbf{S}(\xi) = (\mathbf{E_S},\pi|\mathbf{E_S},\mathbf{X})$ are called the *disk bundle* and *sphere bundle* associated to $\xi$, respectively.

Clearly, the disk and sphere bundles associated to an n-dimensional vector bundle have fibers homeomorphic to $\mathbf{D}^n$ and $\mathbf{S}^{n-1}$, respectively.

Finally, we define what it means for a vector bundle to be oriented.

**Definition.** Let $\xi = (\mathbf{E},\pi,\mathbf{B})$ be a vector bundle. Let $\sigma$ be a map that associates to each $\mathbf{b} \in \mathbf{B}$ an orientation of the vector space $\pi^{-1}(\mathbf{b})$. Such a choice is said to be a *continuously varying* choice of orientations if for every local coordinate chart $(\varphi_\mathbf{b},\mathbf{U_b})$ for $\xi$ and all $\mathbf{b}' \in \mathbf{U_b}$,

$$\sigma(\mathbf{b}') = (T_{\mathbf{b}'})_*(\sigma(\mathbf{b})),$$

where the vector space isomorphism

$$T_{\mathbf{b}'}: \ \pi^{-1}(\mathbf{b}) \to \pi^{-1}(\mathbf{b}')$$

is defined by

$$T_{\mathbf{b}'}(\mathbf{v}) = \varphi_\mathbf{b}(\mathbf{b}',\mathbf{v}).$$

(In other words, the orientations $\sigma(\mathbf{b}')$ are the orientations induced from $\sigma(\mathbf{b})$ using the isomorphisms $T_{\mathbf{b}'}$. See equation (8.8b).) An *orientation* of a vector bundle is a continuously varying choice of orientations in each fiber. A vector bundle is said to be *orientable* if it admits an orientation. An *oriented vector bundle* is a pair $(\xi,\sigma)$, where $\xi$ is a vector bundle and $\sigma$ is an orientation of $\xi$.

**8.9.8. Examples.** It is easy to see that any trivial vector bundle is orientable. The canonical line bundle $\gamma = (\mathbf{E},\pi,\mathbf{P}^1)$ over $\mathbf{P}^1$ is not orientable because a line bundle is orientable if and only if it is trivial (Exercise 8.9.4).

Other examples pertaining to the orientability of vector bundles can be found in the next section.

**Definition.** Let $(\xi_1,\sigma_1)$ and $(\xi_2,\sigma_2)$ be two oriented n-plane bundles and let $F = (\tilde{f},f): \xi_1 \to \xi_2$ be a bundle map. We say that F is an *orientation-preserving bundle map* if $\tilde{f}$ defines an orientation-preserving vector space isomorphism on each fiber of $\xi_1$. We

say that F is an *orientation-reversing bundle map* if $\tilde{f}$ defines an orientation-reversing vector space isomorphism on each fiber of $\xi_1$.

The pullback of an oriented vector bundle has a natural orientation because fibers get mapped isomorphically onto fibers and so we can simply pull back the orientation with the inverse of the isomorphism. More precisely,

**Definition.** Let $(\xi,\sigma)$ be an oriented n-plane bundle and let $f:\mathbf{B}_1 \to \mathbf{B}(\xi)$ be a map. The *induced orientation* $f^*\sigma$ on $f^*\xi$ is defined as follows: Let $(\tilde{f},f):f^*\xi \to \xi$ be the canonical map with fiber maps $\tilde{f}_F$. If $\mathbf{b}_1 \in \mathbf{B}_1$, then

$$f^*(\sigma)(\mathbf{b}_1) = \tilde{f}_F(\mathbf{b}_1)^{-1}{}_*(\sigma(f(\mathbf{b}_1))).$$

**8.9.9. Theorem.** Any vector bundle $\xi = (\mathbf{E},\pi,\mathbf{B})$ over a simply connected space $\mathbf{B}$ is orientable.

**Proof.** See Figure 8.30. Fix a point $\mathbf{b}_0$ in $\mathbf{B}$. Choose an orientation $\sigma_0$ in $\pi^{-1}(\mathbf{b}_0)$. Let $\mathbf{b} \in \mathbf{B}$. We shall define an orientaton $\sigma$ for $\xi$ so that $\sigma(\mathbf{b}_0) = \sigma_0$. Since $\mathbf{B}$ is path-connected, there is a path $\gamma:\mathbf{I} \to \mathbf{B}$ so that $\gamma(0) = \mathbf{b}_0$ and $\gamma(1) = \mathbf{b}$. By Corollary 8.9.5, $\gamma^*\xi$ is a trivial bundle over $\mathbf{I}$, which means that $\gamma^*\xi$ admits a unique orientation $\eta_\gamma$ so that $\eta_\gamma(0)$ is mapped to $\sigma_0$ by the canonical map from $\gamma^*\xi$ to $\xi$. See Exercise 8.9.5 for the uniqueness part. Let $\sigma(\mathbf{b})$ be the orientation of $\pi^{-1}(\mathbf{b})$ to which $\eta_\gamma(1)$ is mapped by the canonical map from $\gamma^*\xi$ to $\xi$. We need to show that

(1) $\sigma$ is a well-defined map, and
(2) $\sigma$ is a continuously varying choice of orientations in each fiber.

We shall only prove (1) and leave (2) as an exercise for the reader. Suppose that there is another path $\lambda:\mathbf{I} \to \mathbf{B}$ so that $\lambda(0) = \mathbf{b}_0$ and $\lambda(1) = \mathbf{b}$. Again, choose the unique orientation $\eta_\lambda$ for $\lambda^*\xi$ so that $\eta_\lambda(0)$ is mapped to $\sigma_0$ by the canonical map from $\lambda^*\xi$ to $\xi$. We must show that $\eta_\lambda(1)$ maps to $\sigma(\mathbf{b})$. Since $\mathbf{B}$ is simply connected, there is a homotopy

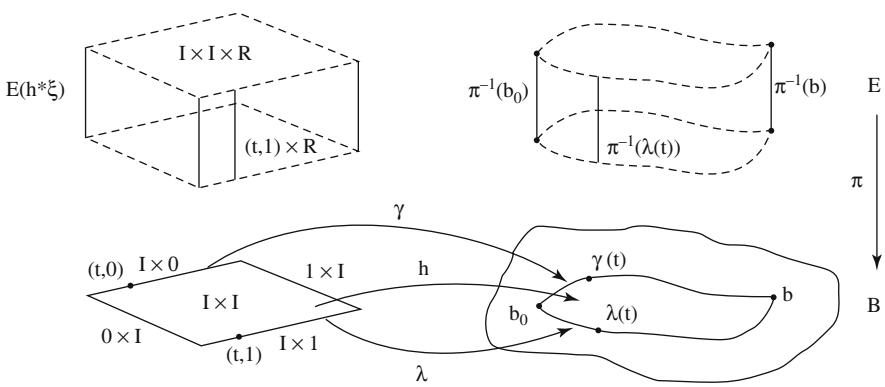



**Figure 8.30.** Proving Theorem 8.9.9.

$$h\colon \mathbf{I}\times\mathbf{I}\to\mathbf{B}$$

between $\gamma(t) = h(t,0)$ and $\lambda(t) = h(t,1)$. Using Corollary 8.9.5 again, it follows that $h^*\varphi$ is a trivial bundle over $\mathbf{I}\times\mathbf{I}$. Identify $\gamma^*\xi$ and $\lambda^*\xi$ with $h^*\xi|\mathbf{I}\times 0$ and $h^*\xi|\mathbf{I}\times 1$, respectively. Now $h^*\xi$ has an orientation $\eta_h$ that agrees with $\eta(0)$ over (0,0). The uniqueness of $\eta_h$ (Exercise 8.9.4) implies that $\eta_h$ agrees with $\eta_\gamma$ and $\eta_\lambda$ over $\mathbf{I}\times 0$ and $\mathbf{I}\times 1$, respectively. Note also that the definition of a homotopy implies $h(1\times\mathbf{I}) = \mathbf{b}$, so that $\eta_h|1\times\mathbf{I}$ must be constant (compare the induced orientation with the given one). In other words, $\eta_h(0,1) = \eta_\gamma(1)$ and $\eta_h(1,1) = \eta_\lambda(1)$ must map to the same orientation of $\pi^{-1}(\mathbf{b})$ and we are done.

## 8.10 The Tangent and Normal Bundles

The last section defined vector bundles and described some of their properties. Vector bundles are very important to the study of manifolds as we shall see in this section.

Let $\mathbf{M}^n$ be a differentiable manifold. Define the n-plane bundle $\tau_{\mathbf{M}} = (\mathbf{E},\pi,\mathbf{M}^n)$ as follows:

$$\mathbf{E} = \bigcup_{\mathbf{p}\in\mathbf{M}^n} T_{\mathbf{p}}(\mathbf{M}^n)$$

and

$$\pi\colon \mathbf{E}\to\mathbf{M}^n \quad \text{sends } \mathbf{v}\in T_{\mathbf{p}}(\mathbf{M}^n) \text{ to } \mathbf{p}.$$

For a coordinate neighborhood $(\mathbf{U},\varphi)$ for $\mathbf{M}$, define

$$\varphi_{\mathbf{U}}\colon \mathbf{U}\times\mathbf{R}^n\to\pi^{-1}(\mathbf{U})$$

by

$$\varphi_{\mathbf{U}}(\mathbf{q},\mathbf{a}) = [\mathbf{U},\varphi,\mathbf{a}]_{\mathbf{q}}.$$

We give $\mathbf{E}$ the weak topology induced by the condition that the sets $\pi^{-1}(\mathbf{U})$ should be open and the maps $\varphi_{\mathbf{U}}$ continuous.

**Definition.** The n-plane bundle $\tau_{\mathbf{M}}$ is called the *tangent bundle* of the manifold $\mathbf{M}^n$.

**8.10.1. Theorem.** Let $\mathbf{M}^n$ be a differentiable manifold and $\tau_{\mathbf{M}} = (\mathbf{E},\pi,\mathbf{M}^n)$ its tangent bundle. Then $\mathbf{E}$ is a 2n-dimensional differentiable manifold and $\pi$ is a differentiable map of rank n.

**Proof.** The proof is straightforward. See [Hirs76]. The natural coordinate neighborhoods of the total space $\mathbf{E}$ are obtained from the compositions of the maps

$$\pi^{-1}(\mathbf{U}) \xrightarrow[\varphi_{\mathbf{U}}^{-1}]{} \mathbf{U}\times\mathbf{R}^n \xrightarrow[\varphi\times\text{identity}]{} \mathbf{R}^n\times\mathbf{R}^n = \mathbf{R}^{2n}.$$

Here are two alternate descriptions of the tangent bundle of a manifold $\mathbf{M}^n$. We shall only define the new total spaces $\mathbf{E}$ in each case because the rest of the definitions are obvious.

**First alternate definition of $\tau_{\mathbf{M}}$:** Let

$$\mathbf{E} = \left( \bigcup_{\text{coordinate neighborhoods}\,(\mathbf{U},\varphi)} \mathbf{U} \times \mathbf{R}^n \right) / \sim \tag{8.28}$$

where the equivalence relation ~ is defined as follows: If $(\mathbf{U},\varphi)\}$ and $(\mathbf{V},\psi)$ are coordinate neighborhoods for $\mathbf{M}^n$, then

$$(\mathbf{q},\mathbf{a}) \in \mathbf{U} \times \mathbf{R}^n$$

is identified with

$$\left(\mathbf{q}, D\left(\psi \circ \varphi^{-1}\right)(\varphi(\mathbf{q}))(\mathbf{a})\right) \in \mathbf{V} \times \mathbf{R}^n \tag{8.29}$$

for all $\mathbf{q} \in \mathbf{U} \cap \mathbf{V}$. Recall our comments about fiber bundles earlier. We can now see how the group GL(n,$\mathbf{R}$) enters the picture, namely, through the linear isomorphism $D(\psi \circ \varphi^{-1})(\varphi(\mathbf{q}))$.

**Second alternate definition of $\tau_{\mathbf{M}}$:** This description is only applicable when $\mathbf{M}^n$ is a submanifold of some $\mathbf{R}^k$: In this case, using our definitions related to the manifolds in Section 8.4, define

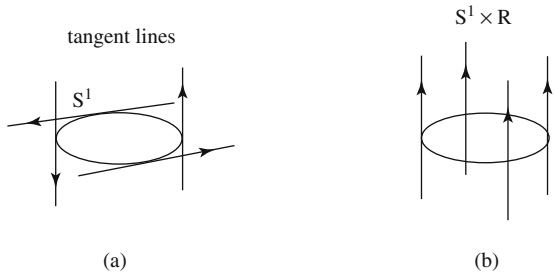
$$\mathbf{E} = \left\{(\mathbf{p},\mathbf{v}) \in \mathbf{M}^n \times \mathbf{R}^k \,\middle|\, \mathbf{v} \text{ is a tangent vector to } \mathbf{M}^n \text{ at } \mathbf{p}\right\}. \tag{8.30}$$

In the terminology of Section 8.4, it could happen that two different points $\mathbf{p}$ and $\mathbf{q}$ of $\mathbf{M}^n$ could call the same $\mathbf{v} \in \mathbf{R}^k$ as their tangent vector, but we differentiated between the uses of $\mathbf{v}$ by the phrase "at $\mathbf{p}$" or "at $\mathbf{q}$." That was an adequate way to deal with the distinction there, but in the context of abstract manifolds it is important that tangent vectors at one point of a manifold are different from those at another point and to have that fact incorporated into the definition. The definition of $\mathbf{E}$ in (8.30) is the most convenient way to accomplish that. The pair $(\mathbf{p},\mathbf{v})$ (as an element of $T_{\mathbf{p}}(\mathbf{M}^n)$) now formalizes the entire phase "the tangent vector $\mathbf{v}$ of $\mathbf{M}^n$ at $\mathbf{p}$." As an added bonus there is no need to worry about what topology to give $\mathbf{E}$ because $\mathbf{E}$ inherits a natural topology as a subspace of $\mathbf{M}^n \times \mathbf{R}^k$. Note that if $\mathbf{M}^n = \mathbf{R}^n$, then the tangent space $T_{\mathbf{p}}(\mathbf{M}^n)$ at a point $\mathbf{p}$ is just $\mathbf{p} \times \mathbf{R}^n$, which agrees with the definition in Section 4.9.

Showing that these two new definitions of $\tau_{\mathbf{M}}$ are equivalent to the original one is left to the reader (Exercise 8.10.1). In the future we shall feel free to choose whichever definition is most convenient.

**8.10.2. Example.** The tangent bundle of $\mathbf{S}^1$ is trivial. We can see that from Figure 8.31. The oriented tangent lines to the circle in $\mathbf{R}^2$ (Figure 8.31(a)) can be rotated into a vertical direction in $\mathbf{R}^3$ (Figure 8.31(b)), so that the total space is nothing but $\mathbf{S}^1 \times$

**Figure 8.31.** The tangent bundle of the circle is trivial.




$\mathbf{R}^1$. A more elegant way to prove this is to define a nonzero cross-section for the tangent bundle and use this to define an isomorphism between the tangent bundle and the trivial bundle.

**Definition.** If the tangent bundle of a manifold is trivial, then the manifold is said to be *parallelizable*.

**8.10.3. Example.** The tangent bundle of $\mathbf{S}^2$ is not trivial. This fact follows from Corollary 8.5.6 using the second alternate definition of the tangent bundle. More generally, since the tangent bundle of **any** even-dimensional sphere $\mathbf{S}^n$ does not admit a cross-section, it is not trivial either.

**Definition.** Let $\mathbf{M}$ be a differentiable manifold with tangent bundle $\tau_{\mathbf{M}}$. A cross-section of $\tau_{\mathbf{M}}$ is called a *vector field* of $\mathbf{M}$. The vector space of vector fields of $\mathbf{M}$ shall be denoted by Vect ($\mathbf{M}$). More generally, if $\mathbf{A} \subseteq \mathbf{M}$, then a cross-section of $\tau_{\mathbf{M}}|\mathbf{A}$ is called *vector field* of $\mathbf{M}$ *defined over* $\mathbf{A}$.

Given our second alternate definition of the tangent bundle of a manifold, it is obvious that the new definition of a vector field for a manifold is equivalent to the definition given in Section 8.5, but now we have a definition that also applies to abstract manifolds.

**Definition.** A *Riemannian metric for a differentiable manifold* is a Riemannian metric for its tangent bundle, which is also assumed to be differentiable if the manifold is. A *Riemannian manifold* is a differentiable manifold together with a Riemannian metric.

By Theorem 8.9.7 we know that every differentiable manifold admits a Riemannian metric, but it is easy to see that directly, because one can always imbed the manifold in some Euclidean space and use the induced inner product on vectors in a tangent plane to the manifold. Because imbeddings are not unique, one can also see from this that many different Riemannian metrics can be defined for a manifold.

**Note.** In future discussions involving differentiable manifolds we shall not hesitate to assume, without any explicit statement, that they have been endowed with a

Riemannian metric if it is convenient. Furthermore, a Riemannian metric on a manifold induces a natural Riemannian metric on every submanifold. We shall always assume that the submanifolds have been given that induced metric.

Real-valued functions on manifolds give rise to vector fields. Assume that a differentiable manifold $\mathbf{M}^n$ has a Riemannian metric <,> and let $f:\mathbf{M} \to \mathbf{R}$ be a differentiable function. Let $\mathbf{p} \in \mathbf{M}$ and, using the functional approach to tangent vectors, let $X_1(\mathbf{p}), X_2(\mathbf{p}), \ldots, X_n(\mathbf{p})$ be an orthonormal basis of the tangent space of $\mathbf{M}$ at $\mathbf{p}$.

**Definition.** The vector field s for $\mathbf{M}$ defined by

$$s(\mathbf{p}) = \sum_{k=1}^{n} X_i(f)X_i(\mathbf{p}).$$

is called the *gradient* of f and is denoted by $\nabla f$ .

The name of the gradient of f is justified because if $\mathbf{M} = \mathbf{R}^n$, then we just get the usual gradient of a function. See Exercise 8.10.2 which also presents an alternative characterization of $\nabla f$.

We return now to the subject of orientation.

**Definition.** An *orientation* of an abstract differentiable manifold $\mathbf{M}$ is an orientation of the tangent bundle of $\mathbf{M}$, that is, a continuously varying choice of orientations of its tangent spaces. A manifold is said to be *orientable* if it admits an orientation. An *oriented manifold* is a pair $(\mathbf{M},\sigma)$, where $\mathbf{M}$ is a manifold and $\sigma$ is an orientation for $\mathbf{M}$.

The notion of orientation for abstract manifolds is compatible with the earlier definition for submanifolds of Euclidean space in Section 8.5. This follows from the fact that Theorem 8.5.3 holds for abstract manifolds, namely, that an n-dimensional closed compact connected abstract manifold is orientable if and only if its nth homology group is isomorphic to $\mathbf{Z}$ (the idea behind the proof is the same).

**Note on orientability.** One problem when discussing orientability of manifolds is that there are different ways to define this concept. This means that one always has to address the issue of compatibility between the different definitions and one ends up having to state several theorems rather than just one. The one general unifying condition for an n-dimensional connected manifold $\mathbf{M}^n$ to be orientable is that $H_n(\mathbf{M},\partial\mathbf{M}) \approx \mathbf{Z}$. One could make that **the** definition of orientability but the question of whether other very useful ways of describing it are compatible would still be there and therefore one would not save oneself any work.

**8.10.4. Example.** To show that the tangent bundle of $\mathbf{S}^n$, $n \geq 1$, is orientable.

**Solution.** We could appeal to the fact that $H_n(\mathbf{S}^n)$ is isomorphic to $\mathbf{Z}$, but it may give the reader a little more understanding if we prove this directly. Let us use the coordinate neighborhoods $(\mathbf{U}_+,\varphi_+)$ and $(\mathbf{U}_-,\varphi_-)$ defined in Example 8.8.4 and the equivalence class of vectors approach to the definition of tangent vectors. If $\mathbf{p} \in \mathbf{U}_+ \cap \mathbf{U}_-$,

then a tangent vector at $\mathbf{p}$ can be represented by a tuple $(\mathbf{U}_+,\varphi_+,\mathbf{a})_\mathbf{p}$ with respect to $(\mathbf{U}_+,\varphi_+)$ or a tuple $(\mathbf{U}_-,\varphi_-,\mathbf{b})_\mathbf{p}$ with respect to $(\mathbf{U}_-,\varphi_-)$, where $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$. The two tuples define the same tangent vector if

$$\mathbf{b} = D\left(\varphi_- \circ \varphi_+{}^{-1}\right)(\varphi_+(\mathbf{p}))(\mathbf{a}).$$

Clearly, the tangent vector representatives $(\mathbf{U}_+,\varphi_+,\mathbf{e}_1)_\mathbf{p}$, $(\mathbf{U}_+,\varphi_+,\mathbf{e}_2)_\mathbf{p}, \ldots, (\mathbf{U}_+,\varphi_+,\mathbf{e}_n)_\mathbf{p}$ define a continuously varying choice of orientations $\sigma(\mathbf{p})$ in the tangent spaces at points $\mathbf{p}$ in $\mathbf{U}_+$. Similarly, we get a continuously varying choice of orientations $\mu(\mathbf{p})$ from the representatives $(\mathbf{U}_-,\varphi_-,\mathbf{e}_1)_\mathbf{p}$, $(\mathbf{U}_-,\varphi_-,\mathbf{e}_2)_\mathbf{p}, \ldots, (\mathbf{U}_-,\varphi_-,\mathbf{e}_n)_\mathbf{p}$. at points $\mathbf{p}$ in $\mathbf{U}_-$. We will get an orientation for the tangent bundle of $\mathbf{S}^n$ from continuously varying orientations $\sigma$ and $\mu$ over $\mathbf{U}_+$ and $\mathbf{U}_-$, respectively, if and only if $\sigma(\mathbf{p}) = \mu(\mathbf{p})$ for all $\mathbf{p} \in \mathbf{U}_+ \cap \mathbf{U}_-$ or $\sigma(\mathbf{p}) \neq \mu(\mathbf{p})$ for all $\mathbf{p} \in \mathbf{U}_+ \cap \mathbf{U}_-$ (we can then simply reverse the orientations of $\mu$). In our case, $\sigma(\mathbf{p}) = \mu(\mathbf{p})$ because by Exercise 8.8.1(b),

$$\left(\varphi_- \circ \varphi_+^{-1}\right)(\mathbf{a}) = \frac{\mathbf{a}}{|\mathbf{a}|^2},$$

so that the linear transformation

$$D\left(\varphi_- \circ \varphi_+^{-1}\right)(\varphi_+(\mathbf{p})):\ \mathbf{R}^n \to \mathbf{R}^n$$

which identifies tangent vector representatives at $\mathbf{p}$ is orientation preserving.

Another interesting fact is that the analog of Theorem 7.5.7 holds for differentiable manifolds. We shall only restate part (1).

**8.10.5. Theorem.** Every simply connected differentiable manifold is orientable.

**Proof.** This follows from Theorem 8.9.9.

Differentiable maps between differentiable manifolds induce bundle maps between their tangent bundles.

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be differentiable manifolds and let $f: \mathbf{M}^n \to \mathbf{N}^k$ be a differentiable map. Define a map

$$Df:\ E(\tau_\mathbf{M}) \to E(\tau_\mathbf{N})$$

by the condition that for all $\mathbf{p} \in \mathbf{M}$,

$$Df \,|\, T_\mathbf{p}(\mathbf{M}) = Df(\mathbf{p}),$$

where

$$Df(\mathbf{p}):\ T_\mathbf{p}(\mathbf{M}) \to T_{f(\mathbf{p})}(\mathbf{M})$$

is the map defined earlier by equations (8.15) or (8.19) depending on how tangent vectors are defined. The vector bundle map

$$F = (Df, f): \ \tau_{\mathbf{M}} \to \tau_{\mathbf{N}}$$

is called the vector bundle map of the tangent bundles *induced* by f.

**8.10.6. Theorem.**

(1) Df is a well-defined differentiable map from $\mathbf{E}(\tau_{\mathbf{M}}) \to \mathbf{E}(\tau_{\mathbf{N}})$.
(2) If $f: \mathbf{M} \to \mathbf{M}$ is the identity map, then Df is the identity map on $\mathbf{E}(\tau_{\mathbf{M}})$.
(2) Let $f: \mathbf{M} \to \mathbf{N}$ and $g: \mathbf{N} \to \mathbf{W}$ be differentiable maps between differentiable manifolds. If

$$h = g \circ f: \ \mathbf{M} \to \mathbf{W},$$

then $Dh = Dg \circ Df$.

**Proof.** The proof of these facts is straightforward using the corresponding properties of differentiable maps between Euclidean spaces.

**Definition.** Let $(\mathbf{M}^n, \sigma)$ and $(\mathbf{N}^n, \tau)$ be two oriented n-dimensional manifolds. A map $f: \mathbf{M} \to \mathbf{N}$ is said to be *orientation preserving* or *reversing* if the induced map between the tangent bundles is orientation preserving or reversing, respectively.

Because Theorem 8.5.3 holds for connected compact closed orientable abstract manifolds, we can define the degree of a map $f: \mathbf{M} \to \mathbf{M}$ for such manifolds $\mathbf{M}$ just like before.

**8.10.7. Theorem.** Let $\mathbf{M}$ be a connected compact closed orientable differentiable manifold and let $f: \mathbf{M} \to \mathbf{M}$ be a diffeomorphism. Then f is orientation preserving if and only if $\deg f = +1$.

**Proof.** The proof involves relating what f does to the top-level homology group to what f does to the fibers of the tangent bundle. One can do this using the definition of the fundamental homology class in the proof of Theorem 8.5.3.

**8.10.8. Example.** Let $n \geq 1$. The reflection

$$r: \ \mathbf{S}^n \to \mathbf{S}^n$$

defined by

$$r(x_1, x_2, \ldots, x_{n+1}) = (-x_1, x_2, \ldots, x_{n+1})$$

is orientation reversing since its degree is $-1$. Because the reader may find it helpful, we shall also work through part of the tangent bundle definition of orientability using equations (8.15) and the coordinate neighborhood $(\mathbf{U}_+, p_n)$, where $\mathbf{U}_+ = \mathbf{S}^n - \{\mathbf{e}_{n+1}\}$ and $p_n$ is the stereographic projection. If $\mathbf{p} \in \mathbf{S}^n$ and $\mathbf{p} \neq \mathbf{e}_{n+1}$, then equations (8.15) imply that

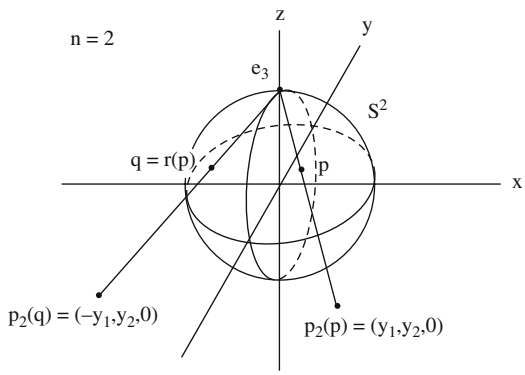



**Figure 8.32.** The case n = 2 in Example 8.10.8.

$$\mathrm{Dr}(\mathbf{p})\left([\mathbf{U}_+, p_n, \mathbf{a}]_{\mathbf{p}}\right) = [\mathbf{U}_+, \mathbf{p}_n, \mathbf{b}]_{r(\mathbf{p})}$$

where

$$\mathbf{b} = D\left(p_n \circ r \circ p_n^{-1}\right)(p_n(\mathbf{p}))(\mathbf{a}),$$

for $\mathbf{a} \in \mathbf{R}^n$. It is easy to check that

$$\left(p_n \circ r \circ p_n^{-1}\right)(x_1, x_2, \ldots, x_n) = (-x_1, x_2, \ldots, x_n).$$

(One can see this intuitively by tracing the action of these maps one after another geometrically. See Figure 8.32. See also Exercise 3.10.1.) This is an orientation reversing linear map and implies that $\mathrm{Dr}(\mathbf{p})$ is orientation reversing on tangent spaces (use an argument similar to the one in Example 8.10.4).

We finish this section with some more important vector bundle related facts for differentiable manifolds.

**Definition.** Let $\mathbf{M}^n$ be a submanifold of a manifold $\mathbf{N}^k$. The *normal bundle* of $\mathbf{M}^n$ in $\mathbf{N}^k$ is the (k − n)-plane vector bundle $\nu_{\mathbf{M}} = (\mathbf{E}, \pi, \mathbf{M}^n)$ defined as follows: Choose a Riemannian metric for $\mathbf{N}^k$ if it does not already have one. Using the natural inclusions and commutative diagram

$$\begin{array}{ccc} \mathbf{E}(\tau_{\mathbf{M}}) & \subseteq & \mathbf{E}(\tau_{\mathbf{N}}) \\ \pi_{\mathbf{M}}\downarrow & & \downarrow\pi_{\mathbf{N}} \\ \mathbf{M} & \subseteq & \mathbf{N} \end{array}$$

for the tangent bundles of $\mathbf{M}$ and $\mathbf{N}$, respectively, we define the fiber of $\nu_{\mathbf{M}}$ over $\mathbf{p} \in \mathbf{M}$ to be the orthogonal complement of the n-dimensional plane $\pi_{\mathbf{M}}^{-1}(\mathbf{p})$ in the k-dimensional vector space $\pi_{\mathbf{N}}^{-1}(\mathbf{p})$ and let $\mathbf{E}$ be the union of these fibers, that is,

$$\mathbf{E} = \bigcup_{\mathbf{p}\in\mathbf{M}} \left\{ \mathbf{v} \in \pi_{\mathbf{N}}^{-1}(\mathbf{p}) \,\middle|\, \mathbf{v} \text{ is orthogonal to } \pi_{\mathbf{M}}^{-1}(\mathbf{p}) \right\} \subseteq \mathbf{E}(\tau_{\mathbf{N}}).$$

Set $\pi = \pi_{\mathbf{N}}|\mathbf{E}$.

**8.10.9. Lemma.** $\nu_{\mathbf{M}}$ is a (k – n)-plane bundle and $\tau_{\mathbf{M}} \oplus \nu_{\mathbf{M}} \approx \tau_{\mathbf{N}}|\mathbf{M}$.

**Proof.** This is immediate from the definitions.

If a different Riemannian metric is chosen, then we will get different but isomorphic normal bundles.

**8.10.10. Theorem.** For every differentiable manifold $\mathbf{M}$ there is a vector bundle $\xi$ over $\mathbf{M}$ so that $\tau_{\mathbf{M}} \oplus \xi$ is trivial.

**Proof.** By Theorem 8.8.7 $\mathbf{M}$ can be imbedded in some Euclidean space $\mathbf{R}^m$. Let $\nu_{\mathbf{M}}$ be the normal bundle of $\mathbf{M}$ in $\mathbf{R}^m$. Since the tangent bundle of $\mathbf{R}^m$ is trivial, Lemma 8.10.9 implies that $\tau_{\mathbf{M}} \oplus \nu_{\mathbf{M}}$ is trivial.

An interesting fact is the following:

**8.10.11. Theorem.** Let $\mathbf{M}^n$ be a differentiable manifold and let $d: \mathbf{M} \to \mathbf{M} \times \mathbf{M}$ be the diagonal map imbedding defined by $d(\mathbf{p}) = (\mathbf{p},\mathbf{p})$. Then the tangent bundle $\tau_{\mathbf{M}}$ is isomorphic to the normal bundle $\nu_{\mathbf{M}}$ of $d(\mathbf{M})$ in $\mathbf{M} \times \mathbf{M}$.

**Proof.** We shall identify $\mathbf{M}$ with $d(\mathbf{M})$. Note first of all that there is a canonical diffeomorphism

$$h\colon\ E(\tau_{\mathbf{M}\times\mathbf{M}}) \to E(\tau_{\mathbf{M}}) \times E(\tau_{\mathbf{M}})$$

defined by

$$h\big([\mathbf{U}_i \times \mathbf{U}_j, \varphi_i \times \varphi_j, \mathbf{v}]_{(\mathbf{p},\mathbf{q})}\big) = \big([\mathbf{U}_i, \varphi_i, \pi_1(\mathbf{v})]_{\mathbf{p}}, [\mathbf{U}_j, \varphi_j, \pi_2(\mathbf{v})]_{\mathbf{q}}\big),$$

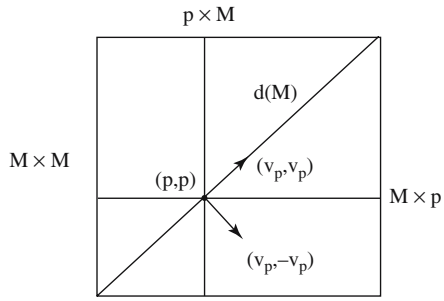
where $(\mathbf{U}_i,\varphi_i)$ and $(\mathbf{U}_j,\varphi_j)$ are coordinate neighborhoods of points $\mathbf{p}$ and $\mathbf{q}$ in $\mathbf{M}$ and the maps

$$\pi_1 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n \quad \text{and} \quad \pi_2 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$$

are the projections onto the first and second factor, respectively (Exercise 8.10.3). Under the identification h, every tangent vector $\mathbf{v}_{(\mathbf{p},\mathbf{q})}$ at $(\mathbf{p},\mathbf{q}) \in \mathbf{M} \times \mathbf{M}$ can be expressed in the form $(\mathbf{v}_{\mathbf{p}},\mathbf{v}_{\mathbf{q}})$, where $\mathbf{v}_{\mathbf{p}}$ is a tangent vector to $\mathbf{M}$ at $\mathbf{p}$ and $\mathbf{v}_{\mathbf{q}}$ is a tangent vector to $\mathbf{M}$ at $\mathbf{q}$. If $\mathbf{M} \times \mathbf{M}$ has been given a Riemannian metric, then $(\mathbf{v}_{\mathbf{p}},\mathbf{w}_{\mathbf{p}})$ will be a tangent vector to $d(\mathbf{M})$ at $(\mathbf{p},\mathbf{p})$ in $\mathbf{M} \times \mathbf{M}$ if and only if $\mathbf{v}_{\mathbf{p}} = \mathbf{w}_{\mathbf{p}}$ and $(\mathbf{v}_{\mathbf{p}},\mathbf{w}_{\mathbf{p}})$ will be a normal vector to $d(\mathbf{M})$ in $\mathbf{M} \times \mathbf{M}$ if and only if $\mathbf{w}_{\mathbf{p}} = -\mathbf{v}_{\mathbf{p}}$. See Figure 8.33. This means that the map

$$\mathbf{v}_{\mathbf{p}} \to (\mathbf{v}_{\mathbf{p}}, -\mathbf{v}_{\mathbf{p}})$$

**Figure 8.33.** Identifying the tangent bundle of **M** with its normal bundle in **M** × **M**.



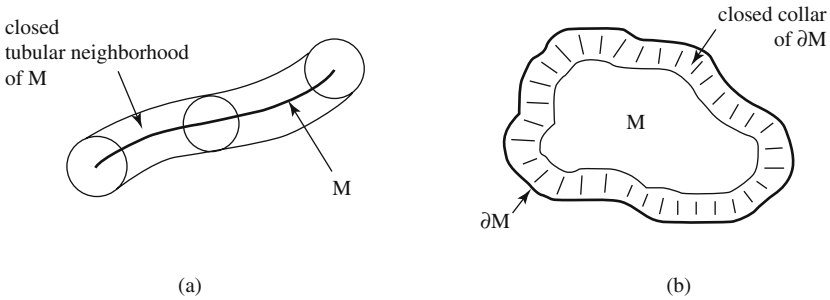




**Figure 8.34.** Tubular neighborhoods and collars.

induces a canonical isomorphism from $\tau_{\mathbf{M}}$ to $\nu_{\mathbf{M}}$.

**Definition.** Let $\mathbf{M}^n$ be a submanifold of a manifold $\mathbf{N}^k$. Suppose that there is a vector bundle $\xi = (\mathbf{E},\pi,\mathbf{M})$ and an imbedding $f:\mathbf{E} \to \mathbf{N}$ so that $f \circ s:\mathbf{M} \to \mathbf{N}$ is the identity map on **M**, where s is the zero section of $\xi$. If f(**E**) is an open neighborhood of **M** in **N**, then the pair $(\xi,f)$ is called a *tubular neighborhood* of **M** in **N**. The associated disk bundle for $\xi$ and the restriction of f to that is called a *closed tubular neighborhood* of **M** in **N**. Often one identifies the total spaces with their image in **N**, so that the subspaces f(**E**) and f(**D**($\xi$)) are also called tubular neighborhoods of **M**, but the bundle structure of the subspaces are assumed to be given in any case.

Figure 8.34(a) shows an example of a closed tubular neighborhood. The total space of the closed tubular neighborhood **D**($\xi$) of $\mathbf{M}^n$ in $\mathbf{N}^k$ is a k-dimensional differentiable manifold with boundary the total space of the associated sphere bundle **S**($\xi$)).

**8.10.12. Theorem.** Every closed submanifold of a closed manifold has a tubular neighborhood.

**Proof.** See [Hirs76].

For more on tubular neighborhoods see [Hirs76]. A related notion is

**Definition.** Let $\mathbf{M}$ be a manifold with boundary $\partial\mathbf{M}$. A *collar* for $\partial\mathbf{M}$ is an imbedding

$$f\colon \partial\mathbf{M} \times [0,\infty) \to \mathbf{M}$$

such that $f(\mathbf{p},0) = \mathbf{p}$ for all $\mathbf{p} \in \partial\mathbf{M}$. The map $f|(\partial\mathbf{M} \times [0,1])$ is called a *closed collar* for $\partial\mathbf{M}$. Again, the subsets of $\mathbf{M}$ that are the images of the maps are sometimes called a collar and closed collar, respectively, but one still assumes that the product structure via a map f is known.

Figure 8.34(b) shows an example of a closed collar.

**8.10.13. Theorem.** Every boundary of a manifold has a collar.

**Proof.** See [Hirs76].

We shall use the last theorem to finish this section by showing that if $(\mathbf{M}^n,\sigma)$ is an oriented n-manifold with boundary, then the orientation $\sigma$ on $\mathbf{M}$ induces a unique orientation $\mu$ on the boundary $\partial\mathbf{M}$. Let $\mathbf{p} \in \partial\mathbf{M}$ and consider $T_{\mathbf{p}}(\partial\mathbf{M})$ as a subset of $T_{\mathbf{p}}(\mathbf{M})$. Let $\mathbf{n_p}$ be a nonzero tangent vector for $\mathbf{M}$ that points into $\mathbf{M}$ and choose a basis $\mathbf{v}_1$, $\mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ for $T_{\mathbf{p}}(\partial\mathbf{M})$ so that $[\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_{n-1},\mathbf{n_p}]$ defines the same orientation of $T_{\mathbf{p}}(\mathbf{M})$ as $\sigma(\mathbf{p})$. (With a Riemannian metric we could have chosen the vector $\mathbf{n_p}$ to be normal to $\partial\mathbf{M}$.) It is easy to show that the orientation $\mu(\mathbf{p}) = [\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_{n-1}]$ of $T_{\mathbf{p}}(\partial\mathbf{M})$ is well-defined. Finally, the fact that the boundary of $\mathbf{M}$ has a collar allows us to choose the vectors $\mathbf{n_p}$ in such a way that they vary continuously with $\mathbf{p}$, so that the function $\mu$ really does define an orientation of $\partial\mathbf{M}$.

**Definition.** The orientation $\mu$ for $\partial\mathbf{M}$ is called the orientation of $\partial\mathbf{M}$ *induced* by the orientation $\sigma$ of $\mathbf{M}$.

## 8.11 Transversality

If there is one concept that is key in the study of manifolds, it is the concept of transversality. We have not run into it much because so many proofs have been omitted in this chapter. Starting with Theorem 8.6.1, which is essential in understanding the structure of manifolds, one would find over and over again that proofs need functions whose singularities are well-structured. The existence of such functions or the ability to deform a given function into one of that type is what transversality is all about. This section will attempt to give a brief overview of some basic definitions and principal results. The material was not presented earlier because we wanted to discuss it in the context of abstract manifolds.

In Section 4.8 we defined what it means for a set in $\mathbf{R}^n$ to have measure 0. We extend this concept to subsets of manifolds.

**Definition.** Let $\mathbf{M}^n$ be a manifold. A subset $\mathbf{A}$ of $\mathbf{M}$ is said to have *measure zero* if for all coordinate neighborhoods $(\mathbf{U},\varphi)$ of $\mathbf{M}$, $\varphi(\mathbf{U} \cap \mathbf{A})$ has measure zero in $\mathbf{R}^n$.

To show that a subset **A** of a manifold has measure zero, it turns out that we do not need to show that $\varphi(\mathbf{U} \cap \mathbf{A})$ has measure zero for all coordinate neighborhoods $(\mathbf{U},\varphi)$, simply for any collection that covers **A**. Furthermore, since n-rectangles do not have measure zero, a closed subset of a manifold that has measure zero is nowhere dense in the manifold.

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be $C^r$ manifolds and let $f: \mathbf{M}^n \to \mathbf{N}^k$ be a $C^r$ map, $r \geq 1$. A point $\mathbf{p} \in \mathbf{M}$ is called a *critical point* and $f(\mathbf{p})$ is called a *critical value* of f if the rank of f at **p** is less than k. If the rank of f equals k at **p**, then **p** is called a *regular point* of f. The points of $\mathbf{N}^k$ that are not critical values are called *regular values*.

The definition of **p** being a regular point is equivalent to requiring that

$$Df(\mathbf{p}): \ T_{\mathbf{p}}\mathbf{M} \to T_{f(\mathbf{p})}\mathbf{N}$$

is onto. It is a good exercise for the reader to convince him/herself that the definitions here reduce to the definitions given in Section 8.6 when we are dealing with submanifolds of $\mathbf{R}^n$ and to the definitions in Section 4.5 when $\mathbf{M} = \mathbf{R}^n$ and $\mathbf{N} = \mathbf{R}$.

**8.11.1. Theorem.** Let $f: \mathbf{M}^n \to \mathbf{N}^k$ be a differentiable map between differentiable manifolds. Let $\mathbf{p} \in \mathbf{M}$ and $\mathbf{q} = f(\mathbf{p})$. If $\mathbf{q} \in \mathbf{N} - \partial\mathbf{N}$ is a regular value for f and $f|\partial\mathbf{M}$, then $f^{-1}(\mathbf{q})$ is a submanifold of **M** of dimension $n - k$, or equivalently, of co-dimension k.

**Proof.** See Theorem 8.3.7.

**8.11.2. Example.** Consider the map

$$f: \ \mathbf{R}^n \to \mathbf{R}$$

defined by

$$f(x_1, x_2, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2.$$

The derivative of f has rank 1 everywhere except at the origin. Therefore, $f^{-1}(1)$, which is just the unit sphere $\mathbf{S}^n$, is a submanifold of dimension $n - 1$ as guaranteed by Theorem 8.11.1.

**Definition.** Let $\mathbf{M}^n$ and $\mathbf{N}^k$ be $C^r$ manifolds and let $f: \mathbf{M}^n \to \mathbf{N}^k$ be a $C^r$ map, $r \geq 1$. We say that the map f is *transverse* to a submanifold **A** in **N** if for all points $\mathbf{p} \in f^{-1}(\mathbf{A})$ and $\mathbf{a} \in \mathbf{A}$

$$T_{\mathbf{a}}(\mathbf{A}) + Df(\mathbf{p})(T_{\mathbf{p}}) = T_{\mathbf{a}}(\mathbf{N}).$$

Theorem 8.11.1 has the following generalization.

**8.11.3. Theorem.** Let $f: \mathbf{M}^n \to \mathbf{N}^k$ be a differentiable map between differentiable manifolds and $\mathbf{A}^d$ a submanifold of **N**. Assume either that

(1) $\partial\mathbf{A} \subset \partial\mathbf{N}$ and the maps f and $f|\partial\mathbf{M}$ are transverse to $\mathbf{A}$, or
(2) $\mathbf{A} \subset \mathbf{N} - \partial\mathbf{N}$ and the maps f and $f|\partial\mathbf{M}$ are transverse to both $\mathbf{A}$ and $\partial\mathbf{A}$.

Then $f^{-1}(\mathbf{A})$ is a submanifold of $\mathbf{M}$ of dimension $n - (k - d)$, or equivalently, of codimension $k - d$, and $\partial f^{-1}(\mathbf{A}) = f^{-1}(\partial\mathbf{A})$.

**Proof.** See Theorem 8.3.7.

If Theorem 8.11.3 looks complicated, it is because we have to worry about boundaries. If there are no boundaries, then the simple conclusion is that $f^{-1}(\mathbf{A})$ is an $(n - (k - d))$-dimensional submanifold whenever f is transverse to $\mathbf{A}$.

**8.11.4. Example.** The inclusion map $i: \mathbf{S}^2 \subseteq \mathbf{R}^3$ is not transverse to the plane $z = 1$. Note that although $i^{-1}(0,0,1) = (0,0,1)$ is a submanifold of $\mathbf{S}^2$, it has the wrong codimension.

**Definition.** We say that two submanifolds $\mathbf{A}$ and $\mathbf{B}$ of a manifold $\mathbf{M}$ *intersect transversally* if

$$T_{\mathbf{p}}(\mathbf{A}) + T_{\mathbf{p}}(\mathbf{B}) = T_{\mathbf{p}}(\mathbf{M})$$

for all points $\mathbf{p} \in \mathbf{A} \cap \mathbf{B}$.

Note that $\mathbf{A}$ and $\mathbf{B}$ intersect transversally if and only if the inclusion map $\mathbf{A} \subseteq \mathbf{M}$ is transverse to $\mathbf{B}$, so that the definition of submanifolds intersecting transversally can be thought of as a special case of the definition of maps being transverse. Note also that the concept of manifolds intersecting transversally basically generalizes what it means for two planes in $\mathbf{R}^n$ to transverse. See Exercise 1.5.18.

**8.11.5. Example.** The lines $x + y = 0$ and $x - y = 0$ intersect transversally at the origin. The unit circle $\mathbf{S}^1$ does not intersect the line $x = 1$ transversally at $(1,0)$.

If two submanifolds $\mathbf{A}^r$ and $\mathbf{B}^s$ of a manifold $\mathbf{M}^n$ intersect transversally, then their intersection is a $(r + s - n)$-dimensional submanifold. In particular, if a compact r-dimensional submanifold and a compact $(n - r)$-dimensional submanifold intersect transversally in $\mathbf{M}$, then their intersection consists of a finite collection of points.

**8.11.6. Theorem.** (The Morse-Sard Theorem) Let $\mathbf{M}^n$, $\mathbf{N}^k$ be manifolds of dimension n, k, respectively, and let

$$f: \mathbf{M} \to \mathbf{N}$$

be a $C^r$ map. Let $\mathbf{C}$ be the set of critical points of f. If

$$r > \max(0, n - k),$$

then $f(\mathbf{C})$ has measure zero in $\mathbf{N}$. The set of regular values of f is dense in $\mathbf{M}$.

**Proof.** See [Hirs76]. Hirsch also explains why one needs the strange constraint on r.

[Hirs76] gives a number of applications of this theorem and Theorem 8.11.3. For example, one can use it to give a fairly simple proof of the fact that there is no differentiable retraction $f: \mathbf{M} \to \partial\mathbf{M}$ (see Exercise 8.11.1 and compare this result with Theorem 7.5.1.4) and the Brouwer fixed-point theorem that every differentiable map $f: \mathbf{D}^n \to \mathbf{D}^n$ has a fixed point. The standard proofs of the continuous versions of these theorems involve algebraic topology. We shall sketch two applications of Theorem 8.11.3 and the Morse-Sard Theorem. The first results in an alternate definition of the degree of a map and the second deals with intersection numbers, vector fields, and the Euler characteristic. The transversality approach to these results will give us additional insights into the underlying geometry.

Let $(\mathbf{M}^n,\sigma)$ and $(\mathbf{N}^n,\mu)$ be closed compact connected oriented differentiable manifolds and assume that

$$f: \mathbf{M}^n \to \mathbf{N}^n$$

is a differentiable map. By the Morse-Sard theorem, f has regular values. Let $\mathbf{q} \in \mathbf{N}$ be one such regular value for f. By Theorem 8.11.3 and the compactness of $\mathbf{M}$, $f^{-1}(\mathbf{q})$ consists of a finite number of isolated points. Let $\mathbf{p} \in f^{-1}(\mathbf{q})$.

**Definition.** The *degree of f at* $\boldsymbol{p}$, $\deg_{\mathbf{p}} f$, is defined by

$$\begin{aligned} \deg_{\mathbf{p}} f &= +1, \quad \text{if } Df(\mathbf{p}) \text{ maps } T_{\mathbf{p}}(\mathbf{M}) \text{ in an orientation preserving way to } T_{\mathbf{q}}(\mathbf{N}) \\ &= -1, \quad \text{if } Df(\mathbf{p}) \text{ maps } T_{\mathbf{p}}(\mathbf{M}) \text{ in an orientation reversing way to } T_{\mathbf{q}}(\mathbf{N}). \end{aligned}$$

Define the *degree of f over* $\boldsymbol{q}$, $\deg(f,\mathbf{q})$, by

$$\begin{aligned} \deg(f,\mathbf{q}) &= \sum_{\mathbf{p}\in f^{-1}(\mathbf{q})} \deg_{\mathbf{p}} f, \text{ if } f^{-1}(\mathbf{q}) \neq \phi, \\ &= 0, \text{ otherwise.} \end{aligned}$$

Note that since the set $f^{-1}(\mathbf{q})$ is finite, the sum in the definition of $\deg(f,\mathbf{q})$ is finite. Figure 8.35 attempts to show what we are doing in the case of a map $f: \mathbf{S}^1 \to \mathbf{S}^1$. The map f indicated in the figure has $f^{-1}(\mathbf{q}) = \{\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3\}$, so that, with the orientations as shown,

$$\deg(f,\mathbf{q}) = 1 - 1 + 1 = 1.$$

Observe also that if we assume that $f(\mathbf{S}^1)$ is a rubber band and pull it tight, then the two point $\mathbf{p}_2$ and $\mathbf{p}_3$ in the pre-image of $\mathbf{q}$ disappear but the degree over $\mathbf{q}$ of the new map is still 1. This "pulling tight" corresponds to changing the map f by a homotopy and should suggest that homotopies do not change the degree.

**8.11.7. Example.** Define $f: \mathbf{S}^1 \to \mathbf{S}^1$ by $f(\mathbf{z}) = \mathbf{z}^n$, $\mathbf{z} \in \mathbf{C}$. Then $\deg(f,1) = n$ because the n nth roots of unity map to 1 in an orientation-preserving way (Exercise 8.11.2).

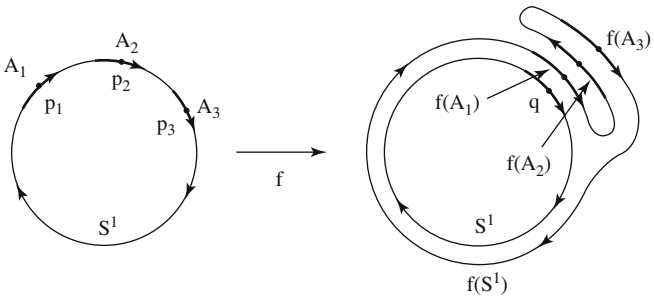



**Figure 8.35.** Making sense of deg(f,**q**).

**8.11.8. Theorem.**

(1) The value deg(f,**q**) does not depend on the choice of regular value **q**.
(2) If f, g: $\mathbf{M} \to \mathbf{N}$ are homotopic differentiable maps and if **q** and **q′** are regular values for f and g, respectively, then deg(f,**q**) = deg(g,**q′**).

**Proof.** See [Miln65b] or [Hirs76].

Theorem 8.11.8 means that the next two definitions are well defined.

**Definition.** The *degree of f*, deg f, is defined to be deg(f,**q**) for any regular value **q** for f.

**Definition.** If g: $\mathbf{M} \to \mathbf{N}$ is any continuous map, then define the *degree of g*, deg g, to be the degree of any differentiable map f: $\mathbf{M} \to \mathbf{N}$ that is homotopic to g.

**8.11.9. Theorem.** The definition of the degree of a map above agrees with the definition of the degree of a map given in Section 7.5.1.

**Proof.** The proof is not hard, but the best way to prove this theorem is to use the tangent bundle of a manifold, its relationship to orientability, and the connection between that and the top-dimensional homology group of the manifold that we discussed in earlier sections.

Next, we use transversality to help shed more light on the duality in manifolds that we discussed in Section 7.5.2. Let $\mathbf{N}^k$ be a closed submanifold of a manifold $\mathbf{W}^{n+k}$. We shall identify $\tau_{\mathbf{N}} \oplus \nu_{\mathbf{N}}$ with $\tau_{\mathbf{W}}|\mathbf{N}$ and assume that the normal n-plane bundle $\nu_{\mathbf{N}}$ is oriented with orientation $\mu$. Let $(\mathbf{M}^n,\sigma)$ be a closed compact oriented manifold. Consider a differentiable map

$$f\colon \mathbf{M}^n \to \mathbf{W}$$

that is transverse to **N**. We know that $f^{-1}(\mathbf{N})$ consists of a finite number of isolated points. Let $\mathbf{p} \in f^{-1}(\mathbf{N})$.

**Definition.** The *index of f at* $\boldsymbol{p}$ *with respect to* $\boldsymbol{N}$, $I_{\mathbf{p}}(f,\mathbf{N})$, is defined by

$$\begin{aligned} \mathbf{I_p} = (f,\mathbf{N}) = +1, &\quad \text{if the composite} \\ &\qquad T_{\mathbf{p}}(\mathbf{M}) \xrightarrow[Df(\mathbf{p})]{} T_{f(\mathbf{p})}(\mathbf{W}) = (\tau_{\mathbf{N}} \oplus \nu_{\mathbf{N}})_{f(\mathbf{p})} \xrightarrow[\text{projection}]{} (\nu_{\mathbf{N}})_{f(\mathbf{p})} \\ &\qquad \text{is orientation preserving (the composite is an isomorphism} \\ &\qquad \text{because of our transversality hypothesis)} \\ = -1, &\quad \text{otherwise.} \end{aligned}$$

The *intersection number of f with respect to* $\boldsymbol{N}$, $I(f,\mathbf{N})$, is defined by

$$\begin{aligned} \mathbf{I}(f,\mathbf{N}) &= \sum_{\mathbf{p} \in f^{-1}(\mathbf{N})} \mathbf{I_p}(f,\mathbf{N}), \quad \text{if } f^{-1}(\mathbf{N}) \neq \boldsymbol{\phi}. \\ &= 0, \qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

**8.11.10. Theorem.** If $f, g: \mathbf{M} \to \mathbf{W}$ are homotopic differentiable maps that are both transverse to $\mathbf{N}$, then $I(f,\mathbf{N}) = I(g,\mathbf{N})$.

**Proof.** See [Hirs76].

By Theorem 8.11.10, the next definition is well defined.

**Definition.** Given an arbitrary continuous map $f: \mathbf{M} \to \mathbf{W}$, let $g: \mathbf{M} \to \mathbf{W}$ be any differentiable map that is homotopic to f and transverse to $\mathbf{N}$. Define the *intersection number of f with respect to* $\boldsymbol{N}$, $I(f,\mathbf{N})$, to be $I(g,\mathbf{N})$.

**Definition.** If $\mathbf{M}^n$ is actually a submanifold of $\mathbf{W}^{n+k}$ and if $i: \mathbf{M} \subseteq \mathbf{W}$ is the inclusion map, then define the *intersection number of* $\boldsymbol{M}^n$ *and* $\boldsymbol{N}^k$ *in* $\boldsymbol{W}^{n+k}$, $I(\mathbf{M},\mathbf{N})$, to be $I(i,\mathbf{N})$. If $\mathbf{M}$ is transverse to $\mathbf{N}$ and if $\mathbf{p} \in \mathbf{M} \cap \mathbf{N}$, then $I_{\mathbf{p}}(i,\mathbf{N})$ is called the *index of the intersection* of $\mathbf{M}$ and $\mathbf{N}$ at $\mathbf{p}$.

The intersection number $I(\mathbf{M},\mathbf{N})$ of the two submanifolds of "dual" dimensions n and k in the (n + k)-dimensional manifold $\mathbf{W}$ is what is most interesting to us here. Compare this with the intersection numbers in Section 7.5.2. The two concepts are closely related. We can carry things a step further to get an intersection number of a single manifold with "itself."

**Definition.** If $\xi$ is an oriented n-plane bundle over a closed compact oriented manifold $\mathbf{M}^n$, then define the *Euler number* of $\xi$, $I(\xi)$, to be $I(s_0(\mathbf{M}),s_0(\mathbf{M}))$, where $s_0$ is the zero cross-section of $\xi$.

Here is a way to visualize the Euler number of the oriented vector bundle $\xi$. Think of having two copies of $\mathbf{M}$ sitting at the zero cross-section in the 2n-dimensional total space of $\xi$. The transversality theorem implies that we can move the second copy of $\mathbf{M}$ slightly so that it meets the first copy transversally. The Euler number is then gotten by assigning a +1 or −1 to each intersection and adding these together. We assign a +1 at an intersection if the orientation of the two copies of $\mathbf{M}$ induce the same orientation as the orientation induced on the total space of $\xi$ by the orientation of $\mathbf{M}$

and $\xi$ and assign a $-1$ otherwise. If $\xi$ is a trivial bundle, the Euler number would be 0 because we can move the second copy of **M** completely off the first so that they do not intersect at all. In fact, all we need to separate the two copies of **M** is a nonzero cross-section, so that we can see that there is a close connection between the Euler number of a vector bundle and the existence of a nonzero cross-section. In general, given a cross-section s the zeros of this cross-section will correspond to where s(**M**) and $s_0$(**M**) will intersect. One can show that the Euler number of $\xi$ can be determined from the intersections of s(**M**) and $s_0$(**M**), no matter which cross-section s one chooses as long as s(**M**) and $s_0$(**M**) meet transversally.

**8.11.11. Theorem.** Let $\xi$ be an oriented n-plane bundle over a closed compact oriented manifold $\mathbf{M}^n$. Then the Euler number of $\xi$ vanishes if and only if $\xi$ admits a nonzero cross-section.

**Proof.** See [Hirs76].

Now it is hard to draw pictures that show what is going on because the dimensions get too large; however, we can show something if we drop the hypotheses that manifolds and bundles are oriented. First, we have to point out that the definitions above dealing with intersection numbers can be given without any assumptions of orientability as long as we do not use signed numbers and work modulo 2, that is, take all values to lie in $\mathbf{Z}_2$ rather than **Z**. This will give us well-defined *mod 2 intersection* and *Euler numbers*. (We could also have defined a *mod 2 degree* of a map between arbitrary manifolds.) Theorem 8.11.11 would hold for the mod 2 Euler number without any hypothesis about orientability.

**8.11.12. Example.** Consider the open Moebius strip line bundle described in Example 8.9.2 and Figure 8.29. Figure 8.36 shows the total space, which is an open Moebius strip. A little thought will convince the reader that it is not possible to move the zero cross-section so that the resulting curve does not intersect the zero cross-section. The best we can do is reduce the number of intersections to one, as is the case with the cross-section $s_1$ in Figure 8.36. Another possible perturbation of the zero cross-section is cross-section $s_2$ which has three intersections but also has a mod 2 Euler number of one. This would imply that the bundle does not have a nonzero cross-section, something we already knew.

An especially interesting case to which Theorem 8.11.11 applies is the tangent bundle $\tau_{\mathbf{M}}$ of **M**. Furthermore, because of Theorem 8.10.11 it follows that the Euler

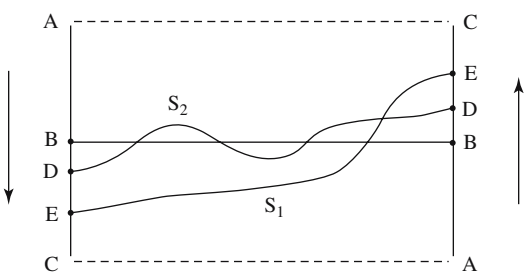



**Figure 8.36.** Case where the mod 2 Euler number equals 1.

number of $\tau_{\mathbf{M}}$ is the intersection number $I(\Delta_{\mathbf{M}},\Delta_{\mathbf{M}})$ of the diagonal $\Delta_{\mathbf{M}}$ with itself in $\mathbf{M} \times \mathbf{M}$.

**8.11.13. Theorem.** The Euler number of the tangent bundle of a closed compact oriented manifold $\mathbf{M}$ is just the Euler characteristic of $\mathbf{M}$, that is, $I(\tau_{\mathbf{M}}) = \chi(\mathbf{M})$.

**Proof.** See [Miln65b] and [MilS74]. Let $s_0$ be the zero cross-section of $\tau_{\mathbf{M}}$. Because the proof of Theorem 8.11.11 would show that $I(\tau_{\mathbf{M}})$ is determined from the intersections of $s(\mathbf{M})$ and $s_0(\mathbf{M})$ for any cross-section s of $\tau_{\mathbf{M}}$, a major part of the proof of this theorem involves finding an s for which one can relate the indices $I_{\mathbf{q}}(s,s_0(\mathbf{M}))$ at a point $\mathbf{q}$ where $s(\mathbf{M})$ and $s_0(\mathbf{M})$ intersect to the Euler characteristic.

Assume that $\mathbf{M}$ has a Riemannian metric $<,>$ and let $f:\mathbf{M} \to \mathbf{R}$ be a function. Consider the gradient vector field $\nabla f$. This vector field vanishes at precisely the critical points of f. Assume that f has only nondegenerate critical points. One can show that at a critical point $\mathbf{p}$ of index k, $I_{s(\mathbf{p})}(s,s_0(\mathbf{M})) = (-1)^k$. This fact and Theorem 8.6.5 proves what we want to show.

Theorem 8.11.13 explains why $I(\xi)$ is called the "Euler" number of the vector bundle $\xi$.

We finish this section by stating one more general result about transversality.

**8.11.14. Theorem.** Let $\mathbf{M}$ and $\mathbf{N}$ be differentiable manifolds and let $\mathbf{A}$ be a submanifold of $\mathbf{N}$. The set of differentiable maps $f:\mathbf{M} \to \mathbf{N}$ that are transverse to $\mathbf{A}$ is dense in the space of all differentiable maps $f:\mathbf{M} \to \mathbf{N}$ (the latter space can be given a natural topology that basically says that functions are close if their derivatives are close).

**Proof.** See [Hirs76].

Theorem 8.11.14 is a really fundamental theorem. It is just one of many theorems of that type. In effect, these theorems say that maps between manifolds are always close to and homotopic to maps that satisfy an appropriate transversality property, so that there is no loss in generality if we assume that the original map has the desired transversality property.

## 8.12 Differential Forms and Integration

In Section 4.8 we defined the integral of real-valued functions defined on subsets of $\mathbf{R}^n$. Section 4.9.1 extended the theory to integrating differential forms on open subsets of $\mathbf{R}^n$. This section will sketch how one can integrate over manifolds. In the case of submanifolds of $\mathbf{R}^n$, given a real-valued function f defined on the manifold, we could of course approximate the manifold by a simplicial complex or other polygonal manifold and then use a Riemann type sum of the form

$$\sum_{\text{polygonal cells } \sigma} \text{volume}(\sigma) f(\mathbf{p}_\sigma),$$

where $\mathbf{p}_\sigma$ is a point of the manifold on or close to the cell $\sigma$, to approximate the integral. This would be the numerical approach. The limit of such sums could be defined to be the integral. Unfortunately, this would not work for abstract manifolds. The approach that works in general and that is also more elegant uses differential forms. We already defined these for open subsets of Euclidean space. To define them for manifolds, we use the fact that manifolds look like Euclidean space locally and use the coordinate neighborhoods of a manifold to carry the definitions we gave for Euclidean space over to the manifold. For a more thorough discussion than given here, see [Spiv70a] or [GuiP74].

**Note.** In this chapter we have given more than one definition for both manifolds and their tangent vectors. Lack of space prevents us from giving all the corresponding variants for definitions in this section. We shall therefore assume that **all manifolds are abstract manifolds** and use the **linear functional approach to tangent vectors** for manifolds. The translation of definition and results for manifolds in $\mathbf{R}^n$ and for the equivalence class of vectors approach will be left as exercises.

Let $\mathbf{M}^n$ be a differentiable manifold (possibly with boundary). Define a vector bundle $\boldsymbol{\omega}^k_{\mathbf{M}} = (\mathbf{E},\pi,\mathbf{M}^n)$ by

$$\mathbf{E} = \bigcup_{\mathbf{p}\in\mathbf{M}^n} \Lambda^k(T_{\mathbf{p}}(\mathbf{M}^n)),$$

and

$$\pi\colon\ \mathbf{E} \to \mathbf{M}^n \text{ is the map that sends } \mathbf{v} \in \Lambda^k(T_{\mathbf{p}}(\mathbf{M}^n)) \text{ to } \mathbf{p}.$$

The local coordinate charts for $\omega^k_{\mathbf{M}}$ as well as the topology for $\mathbf{E}$ are defined in a fashion very similar to what was done for the tangent bundle of $\mathbf{M}$ and we shall leave that as an exercise for the reader. The fibers $\Lambda^k(T_{\mathbf{p}}(\mathbf{M}^n))$ of $\omega^k_{\mathbf{M}}$ are just vector spaces of alternating multilinear maps or exterior k-forms. The total space $\mathbf{E}$ of $\omega^k_{\mathbf{M}}$ is actually a differentiable manifold of dimension $n + \binom{n}{k}$.

**Definition.** The $\binom{n}{k}$-dimensional vector bundle $\omega^k_{\mathbf{M}}$ is called the *exterior k-form bundle* of $\mathbf{M}$.

Other than the definition, the reader does not need to know anything else about $\omega^k_{\mathbf{M}}$. The reason for introducing $\omega^k_{\mathbf{M}}$ is that it provides a convenient way to talk about differential forms below because the best way to think of a differential k-form is as a cross-section in that bundle.

**Definition.** A *differential k-form* on $\mathbf{M}$, or simply *k-form* or *differential form*, is a cross-section of the bundle $\omega^k_{\mathbf{M}}$, that is, it is a map $\omega$ defined on $\mathbf{M}$ that sends $\mathbf{p} \in \mathbf{M}$ to an element $\omega(\mathbf{p}) \in \Lambda^k(T_{\mathbf{p}}(\mathbf{M}))$. The vector space of differential k-forms on $\mathbf{M}$ will be denoted by $\Omega^k{}_0(\mathbf{M})$. There is a *wedge product*

$$\wedge\colon\ \Omega_0^r(\mathbf{M}) \times \Omega_0^s(\mathbf{M}) \to \Omega_0^{r+s}(\mathbf{M})$$

defined by

$$(\omega\wedge\eta)(\mathbf{p}) = \omega(\mathbf{p}) \wedge \eta(\mathbf{p}).$$

Again, a 0-form on **M** is just a real-valued function on **M**. Furthermore, if

$$f\colon\ \mathbf{M} \to \mathbf{R},$$

then Df can really be considered to be a 1-form. The reason is that the tangent bundle for **R** is trivial and there is a canonical identification of all the tangent spaces of **R** with **R**. We capture this idea more precisely with the following definition:

**Definition.** The *differential* of f, denoted by df, is the 1-form on **M** defined by

$$df(\mathbf{p})(\mathbf{v}) = \mathbf{v}(f)$$

for $\mathbf{p} \in \mathbf{M}$ and $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{M})$.

Using coordinate neighborhoods we now relate an arbitrary k-form to those defined on Euclidean space. Let $(\mathbf{U},\varphi)$, $\varphi\colon \mathbf{U} \to \mathbf{R}^n$ be a coordinate neighborhood for **M**. and let

$$\varphi(\mathbf{p}) = (u_1(\mathbf{p}), u_2(\mathbf{p}), \ldots, u_n(\mathbf{p})),$$

where $u_i\colon \mathbf{U} \to \mathbf{R}$. Exercise 8.12.2 asks you to show that the differentials $du_i$ are the dual basis for the tangent vectors $\partial/\partial u_i$. It therefore follows from the properties of the algebra of exterior forms listed in Section 4.9 that every differential k-form $\omega$ on **U** can be written in the form

$$\omega = \sum_{1\le i_1<\ldots<i_k\le n} \omega_{i_1\ldots i_k} du_{i_1} \wedge \ldots \wedge du_{i_k}, \tag{8.31}$$

for functions $\omega_{i_1\ldots i_k}\colon \mathbf{U} \to \mathbf{R}$.

**Definition.** A differential form $\omega$ on **M** is called *continuous*, *differentiable*, *$C^\infty$*, etc., if the functions $\omega_{i_1\ldots i_k}$ in expression (8.31) are continuous, differentiable, $C^\infty$, etc., respectively, with respect to all coordinate neighborhoods $(\mathbf{U},\varphi)$. The vector subspace of $\Omega^k{}_0(\mathbf{M})$ that consists of $C^\infty$ differential k-forms on **M** will be denoted by $\Omega^k(\mathbf{M})$.

It is easy to show that the definitions are well defined and do not depend on any particular coordinate neighborhood. We shall always assume that we have $C^\infty$ manifolds, $C^\infty$ maps, and $C^\infty$ differential forms. Note that, although we used different definitions for the tangent space, this definition of $\Omega^k(\mathbf{R}^n)$, where $\mathbf{R}^n$ is thought of as a manifold, and the one in Section 4.9 agree under the natural correspondence between the tangent space definitions.

**8.12.1. Lemma.** Let $f: \mathbf{M} \to \mathbf{R}$ and let $(\mathbf{U},\varphi)$, $\varphi: \mathbf{U} \to \mathbf{R}^n$ be a coordinate neighborhood for $\mathbf{M}$. If

$$\varphi(\mathbf{p}) = (u_1(\mathbf{p}), u_2(\mathbf{p}), \ldots, u_n(\mathbf{p})),$$

then

$$df = \frac{\partial f}{\partial u_1} du_1 + \frac{\partial f}{\partial u_2} du_2 + \ldots + \frac{\partial f}{\partial u_n} du_n \tag{8.32}$$

on $\mathbf{U}$.

**Proof.** This is a straightforward consequence of the definitions and Theorem 4.9.2.

**Definition.** Given a map $f: \mathbf{M}^n \to \mathbf{N}^m$ between differentiable manifolds, there is a well-defined *induced map*

$$f^*: \Omega^k(\mathbf{N}^m) \to \Omega^k(\mathbf{M}^n) \tag{8.33}$$

that is defined just like the map in equations (4.33) and (4.34) in Section 4.9.

**8.12.2. Theorem.** Let $f: \mathbf{M}^n \to \mathbf{N}^m$ and $g: \mathbf{N}^m \to \mathbf{R}$ be differentiable functions. The map $f^*$ on differentiable forms satisfies

(1) $f^*(\omega_1 + \omega_2) = f^*(\omega_1) + f^*(\omega_2)$
(2) $f^*(g\omega) = (g \circ f)\, f^*\omega$
(3) $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$
(4) Assume that $n = m$ and let $\mathbf{p} \in \mathbf{M}$ and $\mathbf{q} = f(\mathbf{p})$. If $(\mathbf{U},\varphi)$ and $(\mathbf{V},\psi)$ are coordinate neighborhoods for $\mathbf{p}$ in $\mathbf{M}$ and $\mathbf{q}$ in $\mathbf{N}$, respectively, with

$$\varphi(\mathbf{p}) = (u_1(\mathbf{p}), u_2(\mathbf{p}), \ldots, u_n(\mathbf{p})) \text{ and } \psi(\mathbf{q}) = (v_1(\mathbf{q}), v_2(\mathbf{q}), \ldots, v_n(\mathbf{q})),$$

then

$$f^*(g\, dv_1 \wedge dv_2 \wedge \ldots \wedge dv_n) = (g \circ f) \det\left(\frac{\partial (v_i \circ f)}{\partial u_j}\right) du_1 \wedge du_2 \wedge \ldots \wedge du_n. \tag{8.34}$$

**Proof.** The theorem basically follows from Theorem 4.9.3. More details can be found in [Spiv70a].

Now let $\omega$ be a k-form on $\mathbf{M}$. Given any coordinate neighborhood $(\mathbf{U},\varphi)$ for $\mathbf{M}$, express $\omega$ as in equation (8.31) with respect to $(\mathbf{U},\varphi)$. Define a $(k + 1)$-form $d\omega$ over $\mathbf{U}$ by

$$d\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} d\omega_{i_1 \ldots i_k} \wedge du_{i_1} \wedge \ldots \wedge du_{i_k} \tag{8.35a}$$

$$= \sum_{1 \le i_1 < \ldots < i_k \le n} \sum_{j=1}^{n} \frac{\partial \omega_{i_1 \ldots i_k}}{\partial u_j} du_j \wedge du_{i_1} \wedge \ldots \wedge du_{i_k}. \tag{8.35b}$$

**8.12.3. Lemma.** Equation (8.35) defines a well-defined (k + 1)-form $d\omega$ on **M** that is independent of the choice of $(\mathbf{U},\varphi)$.

**Proof.** This is a straightforward computation.

**Definition.** The (k + 1)-form $d\omega$ is called the *differential* of $\omega$. The map $\omega \to d\omega$ is called the *differential operator* d for differential forms on **M**.

**8.12.4. Theorem.**

(1) If $\omega$ and $\eta$ are two k-forms, then $d(\omega + \eta) = d\omega + d\eta$.
(2) If $\omega$ is an r-form and $\eta$ is a t-form, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^{rs}\omega \wedge d\eta.$$

(3) $d(d\omega) = 0$ for any k-form $\omega$, or simply, $d^2 = 0$.
(4) If $f:\mathbf{M}^n \to \mathbf{W}^m$ is a differentiable map and $\omega$ is a k-form on **W**, then $f^*(d\omega) = d(f^*\omega)$.

**Proof.** This is proved just like Theorem 4.9.4.

We are now ready to discuss integration on manifolds. Integrals will be defined by means of differential forms like we did in Section 4.9.1. We shall continue following the presentation given in [Spiv65] and [Spiv70a]. Recall that the theory of integration developed in Section 4.9.1 used cubes $[0,1]^k$. Such spaces are of course not differentiable manifolds since they have "corners." However, these spaces are nice enough so that everything that we did above, such as defining differential forms, induced maps, etc., could have been done for them and the theorems would also remain true, and so we shall treat them as if that had been done.

To begin with, we need to extend the definitions from Section 4.9.1 to the context of manifolds.

**Definition.** A *singular k-cube* in **M** is a $C^\infty$ function $c:[0,1]^k \to \mathbf{M}$. Define the *(i,j)-face* of c to be the singular (n – 1)-cube, $c_{i,j}$, by $c_{i,j} = c \circ I^k_{(i,j)}$.

**Definition.** A formal linear combinations of singular k-cubes in **M** is called a *singular k-chain* and the set of these is denoted by $\Gamma_k(\mathbf{M})$. The boundary of a singular k-chain and the boundary operator

$$\partial:\ \Gamma_k(\mathbf{M}) \to \Gamma_{k-1}(\mathbf{M})$$

is defined just like in equations (4.36) and (4.37).

**Definition.** If $\omega$ is a k-form on **M** and if $c:[0,1]^k \to \mathbf{M}$ is a singular k-cube, then define the *integral of $\omega$ over c* by

$$\int_c \omega = \int_{[0,1]^k} c^*\omega. \tag{8.36a}$$

The *integral of $\omega$ over singular k-chain*

$$c = \sum_i a_i c_i$$

is defined by

$$\int_c \omega = \sum_i a_i \int_{c_i} \omega. \tag{8.36b}$$

Next, we need some lemmas to show that certain quantities are well defined.

**8.12.5. Lemma.** Let $c:[0,1]^n \to \mathbf{R}^n$ be a singular n-cube and let $\omega$ be an n-form on $\mathbf{R}^n$. If c is one-to-one and $\det c' \geq 0$ everywhere, then

$$\int_c \omega = \int_{c([0,1]^n)} f,$$

where $f:\mathbf{R}^n \to \mathbf{R}$ is the unique function with the property that

$$\omega = f\, dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n.$$

**Proof.** This follows immediately from the definitions, Theorem 8.12.2(4), and the change of variable theorem for integrals.

**8.12.6. Lemma.** Let $h:[0,1]^k \to [0,1]^k$ be a one-to-one and onto $C^\infty$ map and assume that $\det h' \geq 0$ everywhere. If c is any singular k-cube and if $\omega$ is any k-form on $\mathbf{M}^n$, then

$$\int_c \omega = \int_{c \circ h} \omega.$$

**Proof.** This is another straightforward computation using Lemma 8.12.5.

**Definition.** Given a singular k-cube c on $\mathbf{M}^n$ and a $C^\infty$ one-to-one and onto map $h:[0,1]^k \to [0,1]^k$ with $\det h' \neq 0$ everywhere, the map $c \circ h$ is called a *reparameterization* of c. The reparameterization is said to be *orientation preserving* if $\det h' > 0$ everywhere and *orientation reversing* if $\det h' < 0$ everywhere.

Lemma 8.12.6 is very important for defining an integral on an abstract manifold. It shows that integrals of differential forms do not change under orientation-preserving reparameterizations. The reason that we have had to develop differential forms in preparation for defining integration is that they transform correctly. If we had defined the integral of a function $f:\mathbf{M}^n \to \mathbf{R}$ over a singular k-cube c on $\mathbf{M}$ by

$$\int_{[0,1]^k} f \circ c$$

then the result in Lemma 8.12.6 would not always hold. By the change of variable theorem, the integral of a reparameterized function is not the same as the integral of

the original function. The formalism of differential forms and how they map is set up the way it is, precisely so as to eliminate these reparameterization factors. Although manifolds look like Euclidean space locally, the exact identification depends on the coordinate neighborhood that we choose and so intrinsic properties of manifolds must be defined so as to be independent of the choice of such neighborhoods.

**8.12.7. Theorem.** (Stokes' Theorem) If $\omega$ is a $(k-1)$-form on a manifold $\mathbf{M}$ and if c is a k-chain on $\mathbf{M}$, then

$$\int_c d\omega = \int_{\partial c} \omega.$$

**Proof.** The proof of this theorem is similar to the one for Theorem 4.9.1.4. See [Spiv65] or [Spiv70a].

Next, we want to describe how one integrates over a whole manifold. Basically, since we know how to integrate over $\mathbf{R}^n$, we know how to integrate over a coordinate neighborhood. The problem is that there are many ways to cover a manifold with coordinate neighborhoods and we must make sure that we make a consistent definition. First of all, we need to revisit the definition of the orientation of a manifold $\mathbf{M}^n$. The definition that we already have says that it is a continuously varying choice of orientations of its tangent spaces. We want to relate this to differential forms.

**8.12.8. Lemma.** Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of an n-dimensional vector space $\mathbf{V}$. Let $\omega \in \Lambda^n(\mathbf{V})$. If

$$\mathbf{w}_i = \sum_{j=1}^{n} a_{ij}\mathbf{v}_j \in \mathbf{V}, \tag{8.37}$$

then

$$\omega(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n) = \det(a_{ij})\, \omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n). \tag{8.38}$$

**Proof.** Define an n-form $\eta \in \Lambda^n(\mathbf{R}^n)$ by

$$\begin{aligned}&\eta\left((a_{11}, a_{12}, \ldots, a_{1n}), (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, (a_{n1}, a_{n2}, \ldots, a_{nn})\right)\\ &\quad = \omega\left(\sum_{j=1}^{n} a_{1j}v_j, \sum_{j=1}^{n} a_{2j}v_j, \ldots, \sum_{j=1}^{n} a_{nj}v_j\right).\end{aligned}$$

By Proposition 4.9.1(3), $\eta = c \det$, where det is the determinant map of $\mathbf{R}^n$ and c is some constant. Since,

$$c = \eta(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = \omega(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n).$$

we are done.

**8.12.9. Lemma.** Let $\omega$ be an n-form on a Riemannian n-manifold $\mathbf{M}^n$. Let $\mathbf{p} \in \mathbf{M}$. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ are orthonormal bases of $T_{\mathbf{p}}(\mathbf{M})$, then

$$\omega(\mathbf{p})(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \pm\, \omega(\mathbf{p})(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n).$$

**Proof.** This follows from Lemma 8.12.8 and the fact that the matrix $(a_{ij})$ in equation (8.37) that relates the bases is an orthogonal matrix whenever the bases are orthonormal bases, so that its determinant is $\pm 1$.

**8.12.10. Lemma.** Let $(\mathbf{M}^n, \sigma)$ be an oriented Riemannian n-manifold. For each $\mathbf{p} \in \mathbf{M}$, choose an orthonormal basis $\mathbf{v}_{1,\mathbf{p}}, \mathbf{v}_{2,\mathbf{p}}, \ldots, \mathbf{v}_{n,\mathbf{p}}$ of $T_{\mathbf{p}}(\mathbf{M})$, so that $\sigma(\mathbf{p}) = [\mathbf{v}_{1,\mathbf{p}}, \mathbf{v}_{2,\mathbf{p}}, \ldots, \mathbf{v}_{n,\mathbf{p}}]$. There is a unique (nonzero) n-form $\omega$ on $\mathbf{M}$ with the property that

$$\omega(\mathbf{p})(\mathbf{v}_{1,\mathbf{p}}, \mathbf{v}_{2,\mathbf{p}}, \ldots, \mathbf{v}_{n,\mathbf{p}}) = 1. \tag{8.39}$$

**Proof.** By Lemma 8.12.8, for any n-form $\omega$, the value $\omega(\mathbf{p})(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is the same for all orthonormal bases $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $T_{\mathbf{p}}(\mathbf{M})$ with $\sigma(\mathbf{p}) = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$. Furthermore, if $\alpha_1, \alpha_2, \ldots, \alpha_n$ is the dual basis for one such basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, then $\alpha_i(\mathbf{v}_j) = \delta_{ij}$ and Property (4.25d) of the wedge product implies that $(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 1$. It follows easily that we can define $\omega$ by the condition that $\omega(\mathbf{p}) = \alpha_{1,\mathbf{p}} \wedge \alpha_{2,\mathbf{p}} \wedge \ldots \wedge \alpha_{n,\mathbf{p}}$, where the $\alpha_{i,\mathbf{p}}$ are a dual basis for the $\mathbf{v}_{i,\mathbf{p}}$. If $(\mathbf{U}, \varphi)$ is a coordinate neighborhood for $\mathbf{M}$, then $\omega$ will have the form $c\, du_1 \wedge du_2 \wedge \ldots \wedge du_n$ over this neighborhood for some constant c. The uniqueness of the n-form $\omega$ follows easily from the fact that $\Lambda^n(T_{\mathbf{p}}(\mathbf{M})) \approx \mathbf{R}$.

**Definition.** Given an oriented Riemannian manifold $\mathbf{M}^n$, the unique n-form $\omega$ defined by equation (8.39) in Lemma 8.12.10 is called the *volume element* of $\mathbf{M}$ and is usually denoted by dV, except in the case of surfaces where one typically writes dA and in the case of curves where one writes ds.

We shall see shortly why the volume element has the name it has.

Lemma 8.12.8 implies that an n-form $\omega$ on $\mathbf{R}^n$ divides the set of ordered bases of $\mathbf{R}^n$ into two subsets in the same way that we defined an orientation of $\mathbf{R}^n$, namely, one set consists of those bases on which it is greater than zero and the other one consists of those on which it is less than zero. Lemma 8.12.10 shows how an orientation defines a nonzero form. This leads to the following observation that gives us another way of thinking about orientation:

**8.12.11. Theorem.** An n-manifold $\mathbf{M}^n$ is orientable if and only if it has a nonzero n-form.

**Proof.** Assume that $\mathbf{M}$ has a nonzero n-form $\omega$. If $\mathbf{p} \in \mathbf{M}$, then choose the orientation of $T_{\mathbf{p}}(\mathbf{M})$ on whose representatives $\omega$ is greater than zero. We leave it to the reader to show that this defines a continuously varying choice of orientations of the tangent planes. Conversely, assume that $\mathbf{M}$ is orientable. Endow $\mathbf{M}$ with a Riemannian metric, choose an orientation $\sigma$, and apply Lemma 8.12.10.

Moving on to the question of how to piece together local definitions of an integral on a manifold in a consistent way, the next lemma gives us exactly what we need.

**8.12.12. Lemma.** Let $(\mathbf{M}^n,\sigma)$ be an oriented n-manifold. Let $c_1$, $c_2:[0,1]^n \to \mathbf{M}$ be singular n-cubes that are extendable to a diffeomorphism on a neighborhood of $[0,1]^n$. Assume that the extensions of $c_1$ and $c_2$ are orientation-preserving diffeomorphisms with respect to the standard orientation on $\mathbf{R}^n$. If $\omega$ is an n-form on $\mathbf{M}$ with the property that

$$\text{support}(\boldsymbol{\omega}) \subseteq c_1\left([0,1]^n\right) \cap c_2\left([0,1]^n\right),$$

then

$$\int_{c_1} \omega = \int_{c_2} \omega.$$

**Proof.** We have done all the hard work in Lemma 8.12.6. We basically want to apply that lemma in the following way:

$$\int_{c_2} \omega = \int_{c_2 \circ \left(c_2^{-1} \circ c_1\right)} \omega = \int_{c_1} \omega.$$

Certainly, $h = c_2^{-1} \circ c_1$ satisfies the hypothesis in the lemma. The only problem is that h is not defined on all of $[0,1]^n$, but this is where the hypothesis about the support of $\omega$ comes in. The proof of Lemma 8.12.6 remains valid in that case.

**Definition.** Let $\mathbf{M}^n$ be an oriented manifold and let $\omega$ be an n-form on $\mathbf{M}$. If c is a singular n-cube which is extendable to a diffeomorphism on a neighborhood of $[0,1]^n$ and if

$$\text{support}(\omega) \subseteq c\left([0,1]^n\right),$$

then define

$$\int_{\mathbf{M}} \omega = \int_c \omega. \tag{8.40}$$

Lemma 8.12.12 implies that

$$\int_{\mathbf{M}} \omega,$$

if it exists, is a well-defined value and does not depend on the choice of c.

Now let $\mathbf{M}^n$ be an oriented manifold. Let $\omega$ be an n-form on $\mathbf{M}$ that has compact support. Theorems 5.8.6 and 5.8.7 and the compactness of the support of $\omega$ imply that we can find a collection of coordinate neighborhoods $(\mathbf{U}_i,\varphi_i)$ of $\mathbf{M}$ satisfying:

(1) $[0,1]^n \subset \varphi_i(\mathbf{U}_i)$
(2) The maps $c_i = \varphi_i^{-1}|[0,1]^n$ are orientation-preserving singular n-cubes for $\mathbf{M}$ with respect to the standard orientation of $\mathbf{R}^n$.
(3) If $\mathbf{V}_i = c_i((0,1)^n)$, then the collection $\{\mathbf{V}_i\}$ is an open cover of $\mathbf{M}$.
(4) Only finitely many $\mathbf{V}_i$ intersect the support of $\omega$.
(5) There exists a partition of unity $\Phi$ subordinate to the cover $\{\mathbf{V}_i\}$.

We use the partition of unity $\Phi$ in (5) to define the integral of $\omega$.

**Definition.** The *integral of $\omega$ over $\boldsymbol{M}$* is defined by

$$\int_{\mathbf{M}} \omega = \sum_{\alpha \in \Phi} \int_{\mathbf{M}} \alpha \cdot \omega. \tag{8.41}$$

**8.12.13. Lemma.** If $\mathbf{M}^n$ is an oriented manifold, then the integral of an n-form $\omega$ on $\mathbf{M}$ with compact support is well defined.

**Proof.** The first observation is that each term

$$\int_{\mathbf{M}} \alpha \cdot \omega$$

is defined because $\alpha \cdot \omega$ is an n-form to which the definition in equation (8.40) applies. Next, the sum on the right-hand side of equation (8.41) is only a finite sum since by condition (4) above all but a finite number of terms are zero. This shows that the right hand side of equation (8.41) makes sense. We need show that it does not depend on the choice of the partition of unity $\Phi$. Assume that $\Psi$ is another partition of unity. Then

$$\sum_{\alpha \in \Phi} \int_{\mathbf{M}} \alpha \cdot \omega = \sum_{\alpha \in \Phi} \left( \sum_{\beta \in \Psi} \int_{\mathbf{M}} \beta \cdot (\alpha \cdot \omega) \right) = \sum_{\beta \in \Psi} \left( \sum_{\alpha \in \Phi} \int_{\mathbf{M}} \alpha \cdot (\beta \cdot \omega) \right) = \sum_{\beta \in \Psi} \int_{\mathbf{M}} \beta \cdot \omega,$$

which proves the independence and finishes the proof of the lemma.

**8.12.14. Proposition.**

(1) Let $\mathbf{M}^n$ be an oriented manifold. If $\omega_1$ and $\omega_2$ are n-forms on $\mathbf{M}$ with compact support and if $a_1, a_2 \in \mathbf{R}$, then

$$\int_{\mathbf{M}} a_1\omega_1 + a_2\omega_2 = a_1 \int_{\mathbf{M}} \omega_1 + a_2 \int_{\mathbf{M}} \omega_2,$$

that is, the integral is linear in the forms.

(2) Let $\mathbf{M}^n$ and $\mathbf{N}^n$ be oriented manifolds and $f: \mathbf{M} \to \mathbf{N}$ an orientation-preserving diffeomorphism. If $\omega$ is an n-form on $\mathbf{N}$ with compact support, then

$$\int_{\mathbf{N}} \omega = \int_{\mathbf{M}} f^* \omega.$$

**Proof.** This is an easy consequence of what has been proved so far.

Now, if $\mathbf{M}^n$ is an oriented n-manifold with boundary, then we know that the orientation on $\mathbf{M}$ induces a natural orientation on $\partial\mathbf{M}$, so that $\partial\mathbf{M}$ becomes an oriented $(n - 1)$-manifold. We always assume that $\partial\mathbf{M}$ has been given that orientation.

**8.12.15. Theorem.** (Stokes' Theorem) Let $\mathbf{M}^n$ be an oriented differentiable manifold with boundary. If $\omega$ is an $(n-1)$-form on $\mathbf{M}$ with compact support, then

$$\int_{\mathbf{M}} d\omega = \int_{\partial \mathbf{M}} \omega.$$

**Proof.** The proof is surprisingly simple using what we know, but we refer the reader to [Spiv70a] for the details.

The following two well-known theorems are trivial consequences of Stokes's theorem.

**8.12.16. Theorem.** (Green's Theorem) Given a surface $\mathbf{S} \subset \mathbf{R}^2$ and differentiable functions

$$f, g\colon \mathbf{S} \to \mathbf{R},$$

then

$$\int_{\partial \mathbf{S}} f\,dx + g\,dy = \iint_{\mathbf{S}} \left( \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dxdy.$$

We assume that the surface has been given the induced orientation from $\mathbf{R}^2$ and its boundary the induced counter-clockwise orientation.

**Proof.** Exercise 8.12.3.

**8.12.17. Theorem.** (The Divergence Theorem) Let $\mathbf{M}^3$ be a differentiable submanifold of $\mathbf{R}^3$ that has boundary. Let $\mathbf{n}(\mathbf{p})$ be the outward-pointing unit normal vector field on $\partial\mathbf{M}$. We assume that $\mathbf{M}$ has been given the induced orientation and Riemannian metric from $\mathbf{R}^3$. Let F be a differentiable vector field on $\mathbf{M}$. Then

$$\int_{\mathbf{M}} \operatorname{div} F\,dV = \int_{\partial \mathbf{M}} \langle F, \mathbf{n} \rangle\,dA.$$

**Proof.** Exercise 8.12.4.

Two more topics related to differential forms on manifold seem worth mentioning. The first has to do with volume and volume elements.

**Definition.** Given an oriented Riemannian manifold $\mathbf{M}^n$, if dV is the volume element of $\mathbf{M}$, then

$$\int_{\mathbf{M}} dV$$

is called the *volume* of $\mathbf{M}$, assuming that the integral exists.

One can show that this definition of volume agrees with the classical definition of lengths of curves, areas of regions in the plane, or volumes of solids. We shall show some of this in Chapter 9.

Now, the only formula for the volume element that we have at this point (see Lemma 8.12.10 and its proof) is that, for a given tangent space, it is the wedge product of the dual basis of an orthonormal basis for that tangent space. It is convenient to have a formula for it in terms of any basis, not just an orthonormal one.

**8.12.18. Lemma.** Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthonormal basis for an n-dimensional vector space $\mathbf{V}$ and $\mathbf{v}_1^*, \mathbf{v}_2^*, \ldots, \mathbf{v}_n^*$ its dual basis. Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ be any basis for $\mathbf{V}$ and $\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_n^*$ its dual basis. If

$$g_{ij} = \mathbf{w}_i \bullet \mathbf{w}_j \quad \text{and} \quad g = \det(g_{ij}), \tag{8.42}$$

then

$$\pm\sqrt{g}\, \mathbf{w}_1^* \wedge \mathbf{w}_2^* \wedge \ldots \wedge \mathbf{w}_n^* = \mathbf{v}_1^* \wedge \mathbf{v}_2^* \wedge \ldots \wedge \mathbf{v}_n^*. \tag{8.43}$$

**Proof.** Suppose that

$$\mathbf{w}_i = \sum_{j=1}^{n} a_{ij}\mathbf{v}_j.$$

It is easy to check that $(g_{ij}) = (a_{ij})\,(a_{ij})^T$ and taking determinants of both sides we get that $g = (\det\,(a_{ij}))^2$. In particular, $g \geq 0$. Since the vector space $\Lambda^n(\mathbf{V})$ has dimension 1, to prove equation (8.43) we only need to show that both sides evaluate to the same value on some nonzero element. It follows from equation (4.25d) that

$$\sqrt{g}(\mathbf{w}_1^* \wedge \mathbf{w}_2^* \wedge \ldots \wedge \mathbf{w}_n^*)(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n) = \sqrt{g}$$

and

$$(\mathbf{v}_1^* \wedge \mathbf{v}_2^* \wedge \ldots \wedge \mathbf{v}_n^*)(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n) = \det(a_{ij}) = \sqrt{g}.$$

The lemma is proved.

The significance of Lemma 8.12.18 is its application to finding volume elements for manifolds.

**8.12.19. Theorem.** Let $\mathbf{M}^n$ be an oriented Riemannian manifold. Let $(\mathbf{U},\varphi)$ be a coordinate neighborhood for $\mathbf{M}$ and define functions

$$g_{ij} = \frac{\partial}{\partial u_i} \bullet \frac{\partial}{\partial u_j} \quad \text{and} \quad g = \det(g_{ij}).$$

Then

$$dV = \pm\sqrt{g}\, du_1 \wedge du_2 \wedge \ldots \wedge du_n.$$

If we choose a coordinate neighborhood $(\mathbf{U},\varphi)$ so that $\varphi$ is orientation preserving, then

$$dV = \sqrt{g}\, du_1 \wedge du_2 \wedge \ldots \wedge du_n.$$

**Proof.** One simply has to use Lemma 8.12.18 on each tangent space $T_{\mathbf{p}}(\mathbf{M})$ for $\mathbf{p} \in \mathbf{U}$. Recall that the vectors $\partial/\partial u_i$ form a basis for the tangent spaces and the differential forms $du_i$ are the dual basis. If $\varphi$ is orientation preserving, the determinant of the matrix $(a_{ij})$ in the proof of Lemma 8.12.18 will be positive.

Theorem 8.12.19 shows that the volume element for an oriented Riemannian manifold is readily determined from any orientation-preserving coordinate neighborhood.

To define the volume we had to integrate and we only showed how to do this over oriented manifolds. What if a manifold is not orientable? Well, we can still integrate something similar to the volume element. Intuitively, this makes sense since volume is gotten by integration and it is natural to think of every manifold as having a "volume" (with respect to a given Riemannian metric, of course, since area, volume, etc. is meaningless without a metric).

**Definition.** Given a manifold $\mathbf{M}^n$, a function $\omega: \mathbf{M} \to \mathbf{R}$ of the form

$$\omega(\mathbf{p}) = |\eta_{\mathbf{p}}| \quad \text{for some} \quad \eta_{\mathbf{p}} \in \Lambda^n(T_{\mathbf{p}}(\mathbf{M})),$$

that is,

$$\omega(\mathbf{p})(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = |\eta_{\mathbf{p}}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)| \geq 0, \quad \text{for } \mathbf{v}_i \in T_{\mathbf{p}}(\mathbf{M}),$$

is called a *volume element* of $\mathbf{M}$.

Note that we are not saying that a volume element is necessarily a differential form. Furthermore, we are setting it up so that it will lead to an unsigned volume. With respect to a coordinate neighborhood $(\mathbf{U},\varphi)$ it will look like

$$\omega = f\,|du_1 \wedge du_2 \wedge \ldots \wedge du_n|, \quad \text{where } f \geq 0.$$

By Theorem 8.12.19,

$$\omega = \sqrt{g}\,|\,du_1 \wedge du_2 \wedge \ldots \wedge du_n|$$

is one such volume element. One can show that it is possible to define an integral

$$\int_{\mathbf{M}} \omega$$

over any manifold $\mathbf{M}$ for any volume element $\omega$ on $\mathbf{M}$ that has compact support. Sections 9.2 and 9.8 will have more to say about volume.

**Example 8.12.20.** Consider a surface $\mathbf{S}$ with a given Riemannian metric and define a volume element by

$$\omega(\mathbf{p})(\mathbf{v}_{\mathbf{p}}, \mathbf{w}_{\mathbf{p}}) = \text{area of parallelogram spanned by } \mathbf{v}_{\mathbf{p}}, \mathbf{w}_{\mathbf{p}} \in T_{\mathbf{p}}(\mathbf{S}). \tag{8.44}$$

If the surface **S** is oriented, then the integral of this volume element will give the usual area of the surface. On the other hand, this volume element is also defined for the Moebius strip, which we know is a nonorientable surface. In this case one can show that the volume element is **not** a differential form, since a nonorientable n-manifold cannot have a nonzero n-form.

The generalization of the volume element defined by equation (8.44) to n-manifolds is discussed in [Spiv70a]. See also Section 9.8.

Finally, there is a direct connection between differential forms and cohomology. Let $\mathbf{M}^n$ be a differentiable manifold.

**Definition.** A differential form $\omega$ on M is called *closed* if $d\omega = 0$. It is called *exact* if $\omega = d\eta$ for some differential form $\eta$.

If we denote the vector space of closed k-forms on **M** by $Z^k_{(d)}(\mathbf{M})$ and the vector space of exact forms by $B^k_{(d)}(\mathbf{M})$, then $B^k_{(d)}(\mathbf{M})$ is a subspace of $Z^k_{(d)}(\mathbf{M})$ because $d^2 = 0$. It follows that we can define

$$H^k_{(d)}(\mathbf{M}) = \frac{Z^k_{(d)}(\mathbf{M})}{B^k_{(d)}(\mathbf{M})}.$$

**Definition.** The vector space $H^k_{(d)}(\mathbf{M})$ is called the *kth de Rham cohomology group or vector space* of **M**.

A very important result is the fact that the de Rham cohomology groups of a differentiable manifold agree with the ordinary cohomology groups with real coefficients (Section 7.3 described cohomology with integer coefficients), that is,

$$H^k_{(d)}(\mathbf{M}) \approx H^k(\mathbf{M}; \mathbf{R}).$$

This shows that the differential forms of a manifold are intimately tied to the topology of the space. Unfortunately, there is no space to expand on that connection here. See [Spiv70a] or [GuiP74].

## 8.13 The Manifold $P^n$

Again, because projective space is such an interesting and important space, we devote a separate section to study its properties as a differentiable manifold.

First of all, we show that $\mathbf{P}^n$ is a differentiable manifold. We need to show all of its points have coordinate neighborhoods. For i = 1, 2, . . . , n + 1, define subsets $\mathbf{U}_i$ in $\mathbf{P}^n$ by

$$\mathbf{U}_i = \{[c_1, c_2, \ldots, c_{n+1}] \mid c_i \neq 0\} \tag{8.45}$$

and

$$\varphi_i : \mathbf{U}_i \to \mathbf{R}^n$$

by

$$\varphi_i([c_1, c_2, \ldots, c_{n+1}]) = \left(\frac{c_1}{c_i}, \ldots, \frac{c_{i-1}}{c_i}, \frac{c_{i+1}}{c_i}, \ldots, \frac{c_{n+1}}{c_i}\right). \tag{8.46}$$

Clearly, the sets $\mathbf{U}_i$ cover $\mathbf{P}^n$ and the maps $\varphi_i$ are well defined, one-to-one, and onto. Note that

$$\varphi_i^{-1}(x_1, \ldots, x_n) = [x_1, \ldots, x_{i-1}, 1, x_i, \ldots, x_n].$$

**Definition.** The map $\varphi_i$ will be called the *ith standard projection of* $\boldsymbol{P}^n$ *onto* $\boldsymbol{R}^n$. The map $\varphi_i^{-1}$ will be called the *ith standard imbedding of* $\boldsymbol{R}^n$ *in* $\boldsymbol{P}^n$.

This extends the terminology introduced in Section 3.5 where we called $\varphi_{n+1}$ the standard projection of $\mathbf{P}^n$ onto $\mathbf{R}^n$ and $\varphi_{n+1}^{-1}$ the standard imbedding of $\mathbf{R}^n$ in $\mathbf{P}^n$.

Next, if

$$\mathbf{U}_{ij} = \mathbf{U}_i \cap \mathbf{U}_j$$

and

$$\varphi_{ji} = \varphi_j \circ \varphi_i^{-1} : \varphi_i(\mathbf{U}_{ij}) \to \varphi_j(\mathbf{U}_{ij}),$$

then

$$\varphi_{ji}(x_1, \ldots, x_n) = \left(\frac{x_1}{x_j}, \ldots, \frac{x_{i-1}}{x_j}, \frac{1}{x_j}, \frac{x_{i+1}}{x_j}, \ldots, \frac{x_n}{x_j}\right)$$

is a $C^\infty$ map. This shows that $\mathbf{P}^n$ is an n-dimensional $C^\infty$ manifold.

**Definition.** Consider n-dimensional projective space $\mathbf{P}^n$ as the quotient space of $\mathbf{S}^n$ where we have identified antipodal points. Let $[\mathbf{p}] \in \mathbf{P}^n$ denote the equivalence class of $\mathbf{p} \in \mathbf{S}^n$. Define the *canonical line bundle* $\gamma_n = (\mathbf{E}, \pi, \mathbf{P}^n)$ over $\mathbf{P}^n$ as follows:

(1) $\mathbf{E} = \{([\mathbf{p}], t\mathbf{p}) \in \mathbf{P}^n \times \mathbf{R}^{n+1} \mid t \in \mathbf{R}\}$.
(2) $\pi([\mathbf{p}], t\mathbf{p}) = [\mathbf{p}]$.
(3) For local coordinate charts, we choose any open set $\tilde{\mathbf{U}} \subset \mathbf{S}^n$ that does not contain any antipodal points and let $\mathbf{U} = \pi(\tilde{\mathbf{U}}) \subset \mathbf{P}^n$. Define a homeomorphism

$$\varphi_{\mathbf{U}} : \mathbf{U} \times \mathbf{R} \to \pi^{-1}(\mathbf{U})$$

by

$$\varphi_{\mathbf{U}}(\mathbf{x}, t) = (x, t\mathbf{q}),$$

where $\mathbf{q} \in \tilde{\mathbf{U}}$ is the unique point so that $\mathbf{x} = [\mathbf{q}]$.

It is clear that $\gamma_n$ is a vector bundle because the sets **U** in (3) form an open cover of $\mathbf{P}^n$.

**8.13.1. Theorem.** The canonical line bundle $\gamma_n$ over $\mathbf{P}^n$ is nontrivial.

**Proof.** The proof is the same as the proof for $\gamma_1$ in Example 8.9.2.

## 8.14 The Grassmann Manifolds

**Definition.** Let $n > 0$ and $k \geq 0$. The *Stiefel manifold* or *Stiefel variety* $V_n(\mathbf{R}^{n+k})$ is defined to be the subspace of $(\mathbf{S}^{n+k-1})^n$ consisting of all n-tuples $(\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n)$, where the $\mathbf{u}_i$ are an orthonormal set of vectors of $\mathbf{R}^{n+k}$. The *Grassmann manifold* or *Grassmann variety* or *Grassmannian* $G_n(\mathbf{R}^{n+k})$ is defined to be the set of n-dimensional linear subspaces of $\mathbf{R}^{n+k}$ with the quotient topology induced by the projection map

$$\pi_n\colon\ V_n\left(\mathbf{R}^{n+k}\right) \to G_n\left(\mathbf{R}^{n+k}\right)$$

defined by

$$\pi_n((\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n)) = \text{plane spanned by the } \mathbf{u}_i.$$

Stiefel and Grassmann manifolds are, like the projective spaces $\mathbf{P}^n$, very important spaces in topology. They have been studied extensively. Here are a few facts about them:

(1) The pair $(V_n(\mathbf{R}^{n+k}),\pi_n)$ is a locally trivial bundle over $G_n(\mathbf{R}^{n+k})$ with all fibers homeomorphic to $\mathbf{O}(n)$.
(2) The map that sends an n-dimensional linear subspace of $\mathbf{R}^{n+k}$ to its k-dimensional orthogonal complement defines a natural homeomorphism between $G_n(\mathbf{R}^{n+k})$ and $G_k(\mathbf{R}^{n+k})$.
(3) There is a canonical n-dimensional vector bundle $\gamma_n(\mathbf{R}^{n+k}) = (\mathbf{E},p,G_n(\mathbf{R}^{n+k}))$ over $G_n(\mathbf{R}^{n+k})$, where

$$\begin{aligned}\mathbf{E} = ((\mathbf{V},\mathbf{v}) \mid \mathbf{V} &\text{ is an n-dimensional linear subspace of } \mathbf{R}^{n+k} \text{ and } \mathbf{v} \in \mathbf{V}\}\\ &\subset G_n(\mathbf{R}^{n+k}) \times \mathbf{R}^{n+k}\end{aligned}$$

and

$$p(\mathbf{V},\mathbf{v}) = \mathbf{V}.$$

(4) $\mathbf{P}^n = G_1(\mathbf{R}^{n+1})$ and $\gamma_n = \gamma_1(\mathbf{R}^{n+1})$. In other words, the Grassmann manifolds can be thought of as generalizations of projective space.
(5) The n-plane bundles $\gamma_n(\mathbf{R}^{n+k})$ play a fundamental role in the classification of vector bundles over a space. Let **B** be a topological space and let $\mathrm{Vect}_n(\mathbf{B})$ denote the isomorphism classes $[\xi]$ of n-plane bundles $\xi$ over **B**. Define a map

$$\varphi\colon \left[\mathbf{B}, G_n\left(\mathbf{R}^{n+k}\right)\right] \to \mathrm{Vect}_n(\mathbf{B})$$

by

$$\varphi([f]) = \left[f^*\left(\gamma_n\left(\mathbf{R}^{n+k}\right)\right)\right].$$

One can show that $\varphi$ is a bijection whenever $\mathbf{B}$ is paracompact and k is large compared with n.

The first four facts are easy to show. It is the relationship of the Grassmann manifolds to the classification of vector bundles in Fact (5) and the consequences of this that are nontrivial and the most significant for algebraic topology. One has a very good understanding of the structure of these manifolds. Unfortunately, there is no space to expand on this here and all we can do is refer the interested reader to [Stee51], [Munk61], [Huse66], or [MilS74] for more information. However, Grassmann manifolds also play a role in algebraic geometry and we shall run into them again in Chapter 10. We finish this section by showing that they are actually manifolds, justifying the name. (Stiefel manifolds are also but we shall leave that as Exercise 8.14.1.)

**8.14.1. Theorem.** The Grassmann manifold $G_n(\mathbf{R}^{n+k})$ is a compact nk-dimensional $C^\infty$ manifold.

**Sketch of proof.** The compactness follows from the fact that it is the continuous image of the Stiefel manifold, which is compact because it is a closed subspace of a compact space. This also shows second countability. One way to prove that $G_n(\mathbf{R}^{n+k})$ is Hausdorff, is to show that one can define a continuous real-valued function on the space that takes on different values at any two given points. Let $\mathbf{V}, \mathbf{W} \in G_n(\mathbf{R}^{n+k})$. Pick a point $\mathbf{p}$ in $\mathbf{R}^{n+k}$ that belongs to the linear subspace $\mathbf{V}$ but not to $\mathbf{W}$. The function

$$f\colon G_n\left(\mathbf{R}^{n+k}\right) \to \mathbf{R}$$

defined by

$$f(\mathbf{X}) = \mathrm{dist}(\mathbf{p}, \mathbf{X})$$

will do the job. It remains to show that the space is locally Euclidean.

Let $\mathbf{V}$ be an n-dimensional linear subspace of $\mathbf{R}^{n+k}$ and let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be a basis for $\mathbf{V}$. We can represent B by means of an $n \times (n + k)$ matrix $M_B$ whose rows are the vectors $\mathbf{v}_i$. Assume that the first n columns of $M_B$ are linearly independent. It is easy to show that out of all the matrices $M_B$ we get as B ranges over all bases of $\mathbf{V}$ there is a unique one $M_{\mathbf{V}}$ of the form

$$M_{\mathbf{V}} = (I \;\; N_{\mathbf{V}}), \tag{8.47}$$

where I is the $n \times n$ identity matrix and $N_{\mathbf{V}}$ is an $n \times k$ matrix. Now let

$$\pi_{\mathbf{V}}\colon \mathbf{R}^{n+k} \to \mathbf{V}$$

be the orthogonal projection of $\mathbf{R}^{n+k}$ onto $\mathbf{V}$ and let $\mathbf{O}$ be the open neighborhood of $\mathbf{V}$ in $G_n(\mathbf{R}^{n+k})$ consisting of all n-dimensional linear subspace $\mathbf{W}$ of $\mathbf{R}^{n+k}$ that $\pi_{\mathbf{V}}$ projects **onto V**. The map

$$\varphi\colon\ \mathbf{O} \to \mathbf{R}^{nk}$$

defined by

$$\varphi(\mathbf{W}) = (n_{11}, n_{12}, \ldots, n_{1k}, n_{21}, n_{22}, \ldots, n_{n1}, n_{n2}, \ldots, n_{nk}),$$

where $(n_{ij})$ is the $n \times k$ matrix $N_{\mathbf{W}}$ as defined in equation (8.47) with respect to $\mathbf{W}$, is a homeomorphism and $(\mathbf{O},\varphi)$ serves as a coordinate neighborhood for $G_n(\mathbf{R}^{n+k})$.

In the general case all that we can assert is that some n columns of $M_B$ will be linearly independent. In this case we can find a unique basis B so that these columns will define an $n \times n$ identity matrix and the remaining k columns form a unique $n \times k$ matrix which we again denote by $N_{\mathbf{V}}$. There will be a corresponding coordinate neighborhood $(\mathbf{O},\varphi)$. It is not hard to show that all these coordinate neighborhood define a $C^\infty$ structure on $G_n(\mathbf{R}^{n+k})$ and we are done with the proof of Theorem 8.14.1.

## 8.15 Exercises

### Section 8.2

8.2.1. Prove that every regular differentiable map $f\colon \mathbf{R} \to \mathbf{R}$ is one-to-one and onto an open interval. Show by example, that a regular differentiable map $f\colon \mathbf{R}^n \to \mathbf{R}^n$, $n > 1$, need not be one-to-one.

### Section 8.4

8.4.1. Redo Example 8.4.5 but use the spherical coordinate parameterization

$$\Phi(\theta,\phi) = (\cos\phi\sin\phi, \sin\theta\sin\phi, \cos\phi),$$

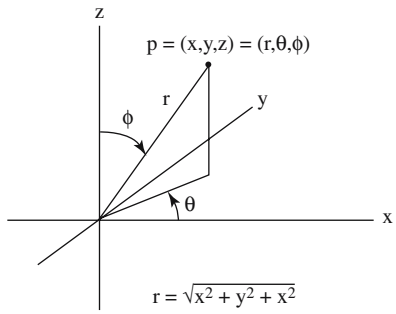
of the sphere, where $(\theta,\phi) \in [0,2\pi] \times [0,\pi]$.

**Note:** The *spherical coordinates* $(r,\theta,\phi)$ of a point $\mathbf{p}$ in $\mathbf{R}^3$ are related to the Cartesian coordinates $(x,y,z)$ of $\mathbf{p}$ via the equations

$$\begin{aligned} x &= r\cos\theta\sin\phi, \\ y &= r\sin\theta\sin\phi, \\ z &= r\cos\phi, \end{aligned}$$

where $r = \sqrt{x^2 + y^2 + z^2}$. See Figure 8.37. In other words, $\theta$ is the polar coordinate angle of the point $(x,y)$ in the plane and $\phi$ is the angle between the z-axis and the vector $\mathbf{p}$. The angle $\theta$ is sometimes called the *azimuth* of $\mathbf{p}$ and the angle $\phi$, the *colatitude* of $\mathbf{p}$. Unfortunately, the notation for spherical coordinates is not as standardized as that for

**Figure 8.37.** Spherical coordinates.




polar coordinates. For example, some books, notably European texts and books on mathematical physics, switch θ and ϕ.

8.4.2. Let **M** be the cylinder in $\mathbf{R}^3$ defined by the equation $x^2 + y^2 = 1$. Find the tangent plane to **M** at the point $\mathbf{p} = (\sqrt{3}, 1, 5)$ in two ways:

(a) Using the cylindrical coordinate parameterization

$$\Phi(\theta, z) = (\cos\theta, \sin\theta, z), \quad (\theta, z) \in [0, 2\pi] \times \mathbf{R}.$$

(b) Using the gradient method like in Example 8.4.7.

**Note:** The *cylindrical coordinates* (r,θ,z) of a point **p** in $\mathbf{R}^3$ are related to the Cartesian coordinates (x,y,z) of **p** via the equations

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \\ z &= z, \end{aligned}$$

where $r = \sqrt{x^2 + y^2}$. In other words, (r,θ) are just the polar coordinates of (x,y) in $\mathbf{R}^2$.

8.4.3. Let **M** be the cylinder in $\mathbf{R}^3$ defined by the equation $x^2 + y^2 = 1$. Define the function $f: \mathbf{M} \to \mathbf{R}$ by $f(x,y,z) = x - y$.

(a) Prove that f is a differentiable function on **M**.
(b) Find $Df(\mathbf{p})((0,1,1)_{\mathbf{p}})$, where $\mathbf{p} = (1,0,0)$.

8.4.4. Fill in the missing details in the proof of Theorem 8.4.9(1).

### Section 8.5

8.5.1. Prove that an isomorphism $T: \mathbf{V} \to \mathbf{W}$ between two k-dimensional vector spaces **V** and **W**, induces a natural one-to-one correspondence $T_*$ between orientations of **V** and **W** defined by

$$T_*([\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]) = [T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_k)],$$

where $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an ordered basis of **V**.

8.5.2. Prove that the concept of continuously varying orientations of the tangent space of a manifold in $\mathbf{R}^n$ does not depend on the choice of parameterizations.

8.5.3. This exercise describes an alternate approach to the orientability of a smooth manifold $\mathbf{M}^k$ in $\mathbf{R}^n$. Given any two local parameterizations $\Phi_{\mathbf{p}}:\mathbf{U}_{\mathbf{p}} \to \mathbf{V}_{\mathbf{p}}$ and $\Phi_{\mathbf{q}}:\mathbf{U}_{\mathbf{q}} \to \mathbf{V}_{\mathbf{q}}$ for $\mathbf{M}$ define

$$d_{\mathbf{p},\mathbf{q}}(\mathbf{r}) = \det\left(\left(\Phi_{\mathbf{q}}^{-1} \circ \Phi_{\mathbf{p}}\right)'\left(\Phi_{\mathbf{p}}^{-1}(\mathbf{r})\right)\right)$$

for $\mathbf{r} \in \mathbf{V}_{\mathbf{p}} \cap \mathbf{V}_{\mathbf{q}}$.

**Definition.** We say that $\mathbf{M}$ is *orientable* if one can choose local parameterizations $\Phi_{\mathbf{p}}$ for $\mathbf{M}$ in such a way that either $d_{\mathbf{p},\mathbf{q}}(\mathbf{r})$ is positive for all $\mathbf{p}$, $\mathbf{q} \in \mathbf{M}$ and $\mathbf{r} \in \mathbf{V}_{\mathbf{p}} \cap \mathbf{V}_{\mathbf{q}}$ or $d_{\mathbf{p},\mathbf{q}}(\mathbf{r})$ is negative for all $\mathbf{p}$, $\mathbf{q} \in \mathbf{M}$ and $\mathbf{r} \in \mathbf{V}_{\mathbf{p}} \cap \mathbf{V}_{\mathbf{q}}$.

Show that this definition of orientability agrees with the one in Section 8.5.

8.5.4. Prove that every simply connected manifold in $\mathbf{R}^n$ is orientable, where the definition of orientable is based on the definition in Exericse 8.5.3.

**Hint:** Pick a point $\mathbf{p}_0$ in the manifold $\mathbf{M}$. For any other point $\mathbf{q} \in \mathbf{M}$ let $\gamma:[0,1] \to \mathbf{M}$ be a path from $\mathbf{p}_0$ to $\mathbf{q}$. Choose a partition $(t_0 = 0, t_1, \ldots, t_k = 1)$ of $[0,1]$ so that, if $\mathbf{p}_i = \gamma(t_i)$, then we have local parameterizations

$$\Phi_{\mathbf{p}_i}:\mathbf{U}_{\mathbf{p}_i} \to \mathbf{V}_{\mathbf{p}_i},\ i = 0,1,\ldots,k,$$

with the property that $\mathbf{V}_{\mathbf{p}_{i-1}} \cap \mathbf{V}_{\mathbf{p}_i} \neq \phi$ for $i = 1, 2, \ldots k$. If any $d_{\mathbf{p}_{i-1},\mathbf{p}_i}(\mathbf{r})$, $\mathbf{r} \in \mathbf{V}_{\mathbf{p}_{i-1}} \cap \mathbf{V}_{\mathbf{p}_i}$, is negative, then replace $\Phi_{\mathbf{p}_i}$ by $\Phi_{\mathbf{p}_i} \circ \alpha$, where $\alpha:\mathbf{U}_{\mathbf{p}_i} \to \mathbf{U}_{\mathbf{p}_i}$ is an orientation-reversing diffeomorphism. Show that these steps lead to a well-defined collection of local parameterizations. The fact that $\mathbf{M}$ is simply connected is needed to show that the choice for $\Phi_{\mathbf{p}}$ does not depend on $\gamma$.

8.5.5. Consider the torus in $\mathbf{R}^3$ which is the surface obtained by rotating the circle in the x-z plane of radius 1 and center (3,0,0) about the z-axis. Define a nonzero normal vector field on this torus.

### Section 8.6

8.6.1. Show that the definition of critical points, critical values, and nondegenerate critical points for functions $f:\mathbf{M} \to \mathbf{R}$ in terms of local coordinates is independent of the choice of local coordinates.

### Section 8.8

8.8.1. (a) Show that the coordinate neighborhoods $(\mathbf{U}_i,\varphi_i)$ in Example 8.8.3 induce a $C^\infty$ structures on $\mathbf{S}^1$.

(b) Consider the coordinate neighborhoods $\{(\mathbf{U}_+,\varphi_+),(\mathbf{U}_-,\varphi_-)\}$ for $\mathbf{S}^n$ that were defined in Example 8.8.4. Show that

$$\varphi_- \circ \varphi_+{}^{-1}(y) = \frac{y}{|y|^2},\ y \in \mathbf{R}^n.$$

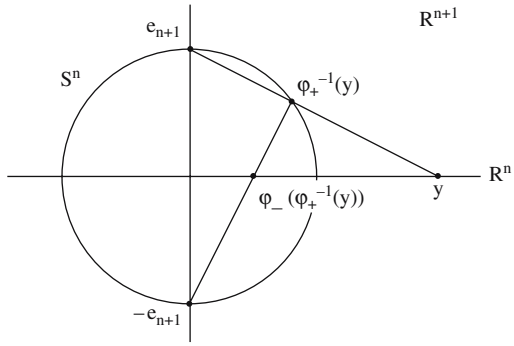



**Figure 8.38.** The stereographic projections of Exercise 8.8.1.

See Figure 8.38. Use this to prove that these coordinate neighborhoods induce a $C^\infty$ structures on $\mathbf{S}^n$.

(c) Show that the $C^\infty$ structures on $\mathbf{S}^1$ defined in (a) and (b) are the same.

8.8.2. Prove the tangent vectors $\mathbf{e}_{i,U}$ defined in the equivalence class of vectors approach to tangent vectors are a basis for $T_{\mathbf{p}}(\mathbf{M})$.

8.8.3. Prove that the tangent vectors $\partial/\partial u_i$ defined in equation (8.16) form a basis of the tangent space.

### Section 8.9

8.9.1. Prove that a trivial n-plane bundles has n linearly independent cross-sections.

8.9.2. Show that the line bundle $\gamma$ defined in Example 8.9.2 is isomorphic to the open Moebius strip line bundle described at the beginning of that example.

8.9.3. Prove the existence of the **continuous** map $\alpha$ in equation (8.27).

8.9.4. Show that a line bundle is orientable if and only if it is trivial.

8.9.5. Show that the orientation of a trivial vector bundle over a path-connected space is uniquely determined once it is specified at one point.

8.9.6. Let $\mathbf{B}$ be a contractible space. Show that every vector bundle over $\mathbf{B}$ is orientable and that the orientation is uniquely determined once it is specified at one point.

8.9.7. Let $\xi = (\mathbf{E},\pi,\mathbf{B})$ be an n-plane bundle. Show that both $\xi \times \xi$ and $\xi \oplus \xi$ are orientable vector bundles.

**Hint:** Show that if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ are bases for the fiber $\pi^{-1}(\mathbf{b})$ over some point $\mathbf{b} \in \mathbf{B}$, then both $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ determine the same orientation of $\pi^{-1}(\mathbf{b}) \times \pi^{-1}(\mathbf{b})$.

### Section 8.10

8.10.1. (a) Show that the alternate description of the tangent bundle of a manifold defined by equations (8.28) and (8.29) produces a vector bundle that is isomorphic to the original definition.

(b) Show the same thing for equation (8.30).

8.10.2. This exercise describes an alternate definition of the gradient of a function $f: \mathbf{M}^n \to \mathbf{R}$. Let $\mathbf{p} \in \mathbf{M}$ and let $(\mathbf{U},\varphi)$ be a coordinate neighborhood of $\mathbf{p}$. If $\varphi(\mathbf{q}) = (u_1(\mathbf{q}),u_2(\mathbf{q}), \ldots ,u_n(\mathbf{q}))$, $\mathbf{q} \in \mathbf{U}$, then define

$$\nabla f(\mathbf{p}) = \sum_{i=1}^{n} \frac{\partial f}{\partial u_i} \frac{\partial}{\partial u_i}.$$

Show that this definition of the gradient of f agrees with the definition in Section 8.10. Show further that this reduces to the standard definition of the gradient of f when $\mathbf{M} = \mathbf{R}^n$.

8.10.3. Prove that the function h in the proof of Theorem 8.10.11 is a diffeomorphism.

8.10.4. Prove that the differentiable manifold $\mathbf{M} \times \mathbf{M}$ is orientable for any differentiable manifold $\mathbf{M}$.

8.10.5. Prove that the total space of the tangent bundle of a differentiable manifold $\mathbf{M}^n$ is always an orientable manifold, even if $\mathbf{M}^n$ is not.

**Section 8.11**

8.11.1. Use transversality to prove that no differentiable retraction $r: \mathbf{M}^n \to \partial\mathbf{M}^n$ exists. (**Hint:** pick a regular value $\mathbf{p} \in \partial\mathbf{M}$ and analyze $\mathbf{N}^{n-1} = r^{-1}(\mathbf{p})$.)

8.11.2. Define $f: \mathbf{S}^1 \to \mathbf{S}^1$ by $f(\mathbf{z}) = \mathbf{z}^n$, $\mathbf{z} \in \mathbf{C}$. Prove that deg $(f,1) = n$.

**Section 8.12**

8.12.1. Let $f: \mathbf{M}^n \to \mathbf{R}$. If we use the equivalence class of vectors approach to defining tangent vectors show that the corresponding definition for the differential of f, df, would be

$$df(\mathbf{p})([\mathbf{U},\varphi,\mathbf{a}]) = D\left(f \circ \varphi^{-1}\right)(\varphi(\mathbf{p}))\mathbf{a},$$

for $\mathbf{p} \in \mathbf{M}$, $(\mathbf{U},\varphi)$ a coordinate neighborhood for $\mathbf{p}$, and $\mathbf{a} \in \mathbf{R}^n$. In particular, show that df is a well-defined element of the dual space of $T_{\mathbf{p}}(\mathbf{M})$.

8.12.2. Let $(\mathbf{U},\varphi)$ be a coordinate neighborhood for a manifold $\mathbf{M}^n$. Let

$$\varphi(\mathbf{p}) = (u_1(\mathbf{p}), u_2(\mathbf{p}), \ldots, u_n(\mathbf{p})),$$

where $u_i: \mathbf{U} \to \mathbf{R}$. Show that $du_1(\mathbf{p})$, $du_2(\mathbf{p})$ , . . . , $du_n(\mathbf{p})$ is the dual basis in $T_{\mathbf{p}}(\mathbf{M})$ for the basis $\partial/\partial u_1$, $\partial/\partial u_2$, . . . , $\partial/\partial u_n$ of tangent vectors.

8.12.3. Prove Green's Theorem, Theorem 8.12.16.

8.12.4. Prove the Divergence Theorem, Theorem 8.12.17.

**Section 8.14**

8.14.1. Prove that the Stiefel manifold $V_n(\mathbf{R}^{n+k})$ is a $C^\infty$ manifold. What is its dimension?

CHAPTER 9

# Differential Geometry

## 9.1 Introduction

This chapter basically continues the study of manifolds, but we now add a metric to the differential structure that we used in the last chapter. We shall assume that we are dealing with a Riemannian $C^\infty$ manifold throughout this chapter. Differential geometry is the study of geometry that deals with metric invariants of such spaces. The differential geometer **can** tell a circle from an ellipse, the topologist cannot. Our goal in this chapter is to discuss those aspects of differential geometry most relevant to computer graphics. Of main concern are smooth curves (one-dimensional manifolds) and surfaces (two-dimensional manifolds). An excellent general reference for someone who wants to get an intuitive feel for the many beautiful geometric insights from which differential geometry developed is [HilC99].

The first half of the chapter deals with metric properties of curves and the second half with metric properties of surfaces. We start in Section 9.2 with the definition and some properties of curve length. Sections 9.3 and 9.4 describe some well-known curvature-related geometric concepts for planar and space curves, respectively. Although planar curves are obviously a special case of space curves, there are useful concepts that make sense for them but not for general curves. Sections 9.6 and 9.7 finish the part on curves by describing some classical constructions. We discuss envelopes of curves, involutes and evolutes, and parallel curves. Next, Section 9.8 begins the study of surfaces. Again, we start with the most basic metric aspect, namely, area and volume. Section 9.9 describes some important curvature-related geometric concepts associated to surfaces. This is a long section, but even so we are only able to describe the most basic of those concepts. Geodesics, or curves of shortest length in a surface, are defined in Section 9.10. Sections 9.11–9.15 discuss a few classical special surfaces, namely, envelopes of surfaces, canal surfaces, involutes and evolutes, parallel surfaces, and ruled surfaces. In Section 9.16 we describe what basically is the modern approach to the material in the first fifteen sections. We finish this chapter with a few comments about possible directions for further study in Section 9.17 and then a summary of important formulas in Sections 9.18 and 9.19.

**Note.** Before we start we need to make something clear. Unless explicitly stated otherwise, we shall assume throughout this chapter that our curves, surfaces, or manifolds are submanifolds of $\mathbf{R}^n$ for some n and that tangent vectors and tangent spaces are defined as in Section 8.4. However, and this is very important to keep in mind, the reader should think of the manifolds as spaces that exist on their own and independent of the engulfing space $\mathbf{R}^n$. The only reason for bringing $\mathbf{R}^n$ into the picture is to simplify the definition of tangent vectors and certain formulas. Everything else will rely on parameterizations, so that the theory would translate to abstract manifolds as defined in Section 8.8 with very little effort.

## 9.2 Curve Length

Chapter 8 already discussed a few facts about curves. Of interest in this chapter are their metric properties and these will be studied by means of parameterizations. We begin with curves in $\mathbf{R}^n$. This will be a warm up for studying curves in manifolds. Defining the length of a curve is the first order of business.

To understand the notion of length, one builds on the everyday meaning of this term in simple cases. To begin with, everyone agrees that the length of a line segment should be the distance between its end points. Next, one wants length to be additive. That leads to defining the length of a polygonal curve as the the sum of the lengths of its segments, which in turn suggests the definition for a general curve.

Let

$$F: [a,b] \to \mathbf{R}^n$$

be an arbitrary parametric curve. The standard approach to defining its length is to approximate the curve with a piecewise linear one and then use the length of the piecewise linear version as an approximation for the length of the curve we are after. More precisely, let $P = (t_0, t_1, \ldots, t_k)$ be a partition of [a,b] and let $\mathbf{p}_i = F(t_i)$. See Figure 9.1.

**Definition.** The *length* of the parametric curve F, denoted by $l_a^b F$, is defined by

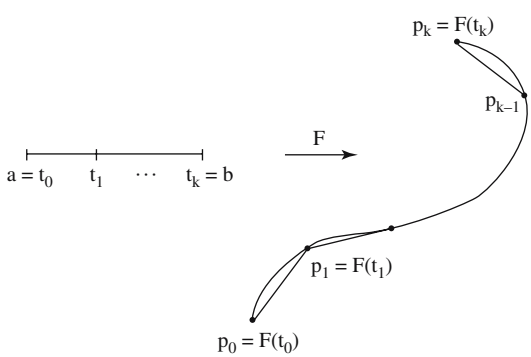



**Figure 9.1.** Curve length is computed from polygonal approximations.

$$l_a^b F = \lim_{\text{partition } P \text{ of } [a,b], |P| \to 0} \sum_{i=1}^{k} |\mathbf{p}_{i-1}\mathbf{p}_i|,$$

provided that this limit exists. ($|P|$ denotes the norm of the partition.) A parametric curve that has a length is called a *rectifiable curve*.

Computing limits would be complicated. Fortunately, we can compute integrals instead.

**9.2.1. Proposition.** If F is $C^1$, then $l_a^b$ F exists and

$$l_a^b F = \int_a^b |F'|.$$

**Proof.** We shall only sketch a proof. For simplicity assume that n = 2. Let F(t) = (x(t),y(t)) and $\mathbf{p}_i = (x_i, y_i) = F(t_i)$. The Proposition basically follows from the mean value theorem, which implies that

$$\begin{aligned} x_i - x_{i-1} &= x'(\alpha)(t_i - t_{i-1}), \\ y_i - y_{i-1} &= y'(\beta)(t_i - t_{i-1}), \end{aligned}$$

for some $\alpha, \beta \in [t_{i-1}, t_i]$. In other words,

$$\begin{aligned} |\mathbf{p}_{i-1}\mathbf{p}_i| &= (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \\ &= (t_i - t_{i-1})^2\left[x'(\alpha)^2 + y'(\beta)^2\right]. \end{aligned}$$

Summing these expressions gives something very much like a Riemann sum that converges to the desired integral.

**9.2.2. Example.** To compute the lengths of

$$\begin{aligned} F(t) &= (t,t), && t \in [0,1], \\ G(t) &= \left(t^2, t^2\right), && t \in [0,1], \quad \text{and} \\ H(t) &= (\sin t, \sin t), && t \in [0, \pi/2]. \end{aligned}$$

**Solution.** By Proposition 9.2.1

$$\begin{aligned} l_0^1 F &= \int_0^1 \sqrt{2} = \sqrt{2}, \\ l_0^1 G &= \int_0^1 2\sqrt{2}t = \sqrt{2}, \quad \text{and} \\ l_0^{\pi/2} H &= \int_0^{\pi/2} \sqrt{2}\cos t = \sqrt{2}. \end{aligned}$$

We should note that although we can now compute lengths using integration, it is still not that easy. The fact that there is a square root under the integral sign means that it is in general pretty much impossible to compute the integral in closed form. However, one can get good approximations using numerical analysis. Another obser-

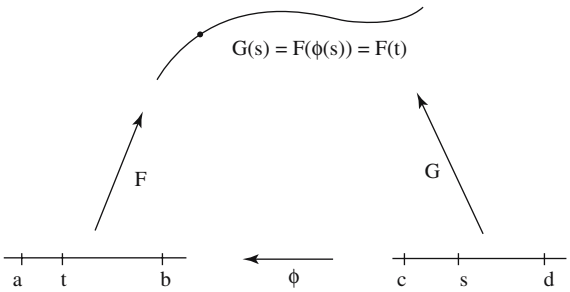



**Figure 9.2.** Curve length is independent of change of parameters.

vation we want to make now will have some important consequences later. The formula in Proposition 9.2.1 shows that the length of a parametric curve is determined by the lengths of its tangent vectors. Tangent vectors lie in ordinary vector spaces and so the essential ingredient in a definition of the length of a curve is a notion of the length of vectors in vector spaces. We shall see in Sections 9.8 and 9.17 that abstractly it is better not to think of length as a generalization of line segment length but rather to think of it more generally as derived from an inner product on vector spaces.

From Example 9.2.2 we see that three different parameterizations of the segment from [0,0] to [1,1] produced the same length. This is of course what we want. We do not want length to be something that depends on the parameterization of a set. To what extent is this true? If we think of the parameter as time and the parametric curve itself as describing a path along which one is walking, then two paths are going to have different lengths if one of them backtracks and the other does not. But if we are interested in the length of a set then we do not want to allow any backtracking anyway and so it is not unreasonable to restrict ourselves to comparing two parametric curves that are both, at least locally, one-to-one functions.

Let $\mathbf{C}$ be the underlying set of two $C^1$ curves $F:[a,b] \to \mathbf{C}$ and $G:[c,d] \to \mathbf{C}$. Let $\phi:[c,d] \to [a,b]$ be a function such that $G(s) = F(\phi(s))$, and let $t = \phi(s)$. See Figure 9.2. The map $\phi$ can be thought of as a change in coordinates. We have the following chain of equalities

$$\begin{aligned}
l_c^d G &= \int_c^d |G'(s)|\, ds \\
&= \int_c^d |F'(\phi(s))\phi'(s)|\, ds \\
&= \int_c^d |F'(\phi(s))|\, \phi'(s)\, ds \quad (\textbf{if } \phi' > 0) \\
&= \int_a^b |F'(t)|\, dt \qquad\qquad \text{(change of variables)} \\
&= l_a^b F.
\end{aligned}$$

This shows what we were trying to show, namely, that the length of a parametric curve is basically an invariant of the underlying set. The next two facts allow us to define the length of this set precisely.

**9.2.3. Lemma.** Let F and G be $C^1$ curves that parameterize the same set **C**. If G(s) = F(ϕ(s)) is a regular reparameterization of **C**, then F and G have the same length.

**Proof.** We leave it to the reader to flesh out the discussion above to a complete proof that handles the case $\phi' < 0$. (We only know that $\phi' \neq 0$.)

**9.2.4. Lemma.** Any compact curve **C** admits a proper regular parameterization $F:[a,b] \to \mathbf{C}$.

**Sketch of Proof.** Since **C** is compact, there are a finite number of local parameterizations that cover **C**. The map F is gotten by piecing together these local parameterizations.

We need compactness in Lemma 9.2.4 because F is supposed to have a compact domain.

**Definition.** Let **C** be a compact curve in $\mathbf{R}^n$. Define the *length* of **C**, denoted by length(**C**), to be the length of a proper regular parameterization of **C**.

Compactness is needed here because, for example, the x-axis is a curve and it clearly does not have finite length.

**9.2.5. Theorem.** The length of a compact curve **C** is well defined.

**Sketch of Proof.** There are two steps involved in the proof. First, one needs to know that proper regular parameterizations F exist. This is Lemma 9.2.4. Next, one has to show that the length does not depend on the choice of F. This is done by showing that any other such map is a regular reparameterization of F and applying Lemma 9.2.3.

We have seen that sets do not have unique parameterizations. However, there is one parameterization for curves that is particularly nice.

**Definition.** If a differentiable curve $F:[a,b] \to \mathbf{R}^n$ ($a \neq b$), satisfies the property that

$$\int_a^t |F'| = t$$

for all $t \in [a,b]$, then F is called the *arc-length parameterization* of the set $\mathbf{C} = F([a,b])$. (Arc-length parameterization is left **undefined** for single-point sets because it is uninteresting and is never used.)

Intuitively, arc-length parameterization has the property that at time t we are a distance t along the curve from the start point. This property also defines arc-length parameterizations of a path uniquely (given a starting point). Furthermore, a simple consequence of the definition is that the lower bound a must be 0 since the integral from a to a is 0. Also, since b is just the length L of the curve, one usually writes the domain of the curve as [0,L]. Another convention is that the parameter s is used rather than t.

**Note.** In the future, **using s** as the parameter of a parametric curve **will mean** that we are dealing with an **arc-length** parameterization.

**9.2.6. Proposition.** Let $F:[0,L] \to \mathbf{R}^n$ be the arc-length parameterization of some path. Then for all $s \in [0,L]$

(1) $|F'(s)| = 1$, and
(2) $F'(s) \bullet F''(s) = 0$.

**Proof.** Property (1) follows from the Fundamental Theorem of Calculus by differentiating the defining equation for arc-length parameterization. Next, we can restate (1) in the form

$$F'(s) \bullet F'(s) = 1.$$

Differentiating both sides of this equation gives (2).

What Proposition 9.2.6 says is that if we are dealing with an arc-length parameterization of a path, then the length of the tangent vector of that parameterization (or its speed) is 1 and the tangent vector is orthogonal to its second derivative. As we go along, we shall see that by using arc-length parameterization we can usually give much cleaner definitions of geometric concepts associated to **paths** (sets that we are typically after and not some accidental parameterization of them) and proofs will be easier. The practical downside is that paths are unfortunately not usually presented via such parameterizations, and so the question arises as to how one might find the arc-length parameterization of a path. We address this question next.

Suppose that $F:[a,b] \to \mathbf{R}^n$ is a curve. The curve F may not be the arc-length parameterization for $\mathbf{C} = F([a,b])$. To find the arc-length parameterization G(s) for $\mathbf{C}$ let us look for a change in coordinate function such that $G(s) = F(\phi(s))$. See Figure 9.2 again. Our earlier discussion then shows that

$$\phi^{-1}(t) = s = \int_0^s |G'| = \int_a^t |F'| \tag{9.1}$$

This means that we know $\phi^{-1}$ and can therefore solve for the function $\phi$ itself.

**9.2.7. Theorem.** Every regular parameterization of a compact curve can be reparameterized into an arc-length parameterization for the curve with an orientation-preserving change of coordinates. The resulting unique arc-length parameterization is called the *induced arc-length parameterization*.

**Proof.** The theorem follows easily from the observations preceding it.

Equation (9.1) leads to two other useful observations, namely,

$$\frac{ds}{dt} = |F'(t)| \tag{9.2}$$

and the following shorthand relation between differential operators

$$\frac{d}{ds} = \frac{1}{|F'(t)|}\frac{d}{dt}. \tag{9.3}$$

Equation (9.3) states an especially handy relationship because it allows us to replace differentiation with respect to arc-length by a differentiation with respect to the actual parameter at hand.

**9.2.8. Example.** To find the arc-length parameterization of the function

$$F(t) = (\sin t, \sin t), \quad t \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right],$$

which parameterizes the segment $[(-1/\sqrt{2}, -1/\sqrt{2}), (1/\sqrt{2}, 1/\sqrt{2})]$.

**Solution.** We have

$$\begin{aligned} s &= \int_{-\pi/4}^{t} |F'(t)|\, dt \\ &= \int_{-\pi/4}^{t} \sqrt{2}\cos t\, dt \\ &= \sqrt{2}\sin t\Big|_{-\pi/4}^{t} \\ &= \sqrt{2}\sin t + 1. \end{aligned}$$

It follows that

$$t = \sin^{-1}\frac{s-1}{\sqrt{2}}.$$

Therefore,

$$G(s) = \left(\sin\left(\sin^{-1}\frac{s-1}{\sqrt{2}}\right), \sin\left(\sin^{-1}\frac{s-1}{\sqrt{2}}\right)\right) = \frac{1}{\sqrt{2}}(s-1, s-1).$$

It is easy to check that this G is an arc-length parameterization.

Example 9.2.8 and the discussion before it show that one can find arc-length parameterizations in a systematic way, although this may have more of a theoretical value than a practical one. First, the square root in the integrand in the integral for arc-length makes that integration difficult in all but the most trivial examples. Second, finding the inverse of a function is rarely easy. Of course, numeric solutions are another matter and quite feasible.

## 9.3 The Geometry of Plane Curves

A natural place at which to start if one wants to understand the geometry of curves is to try to capture the concept of "curvedness." We begin with planar curves. So what exactly should we mean by the curvature of our curve at some point? Just as, for example, in the case of length where one started by agreeing on what one meant by

the length of a simple curve like a segment, we must agree on what curvature should mean in "obvious" cases. It is natural to take the circle as a starting point for curvature. The first assumption we make is that a circle should be said to have constant curvature. Furthermore, small circles should be considered more "curved" than large ones. We can capture these aspects of circles by having the curvature of a circle of radius r be a number proportional to 1/r and nothing will be lost if we normalize things and define the curvature to **be** 1/r. Next, however we finally define curvature for arbitrary curves, if we apply it to a circle, the definition should agree with the one we just gave. Before we present a specific general definition of curvature we look at two possible geometric approaches to getting such a definition that will have the properties we want.

Let

$$F: [a,b] \to \mathbf{R}^2$$

be the parameterization of some curve in the plane. Consider a fixed point $F(t_1)$ on the curve.

**First geometric definition of curvature:** The idea here is to define the curvature at the point $F(t_1)$ to be the reciprocal of the radius of the best "matching" circle to the curve at that point. To find this circle, let $\mathbf{C}(t_1,t_2,t_3)$ denote the circle through the points $F(t_1)$ and two neighboring points $F(t_2)$ and $F(t_3)$. See Figure 9.3. This circle will exist and be unique as long as the points $F(t_i)$ are not collinear. We now let $\mathbf{C}$ be the limit of these circles as $t_2$ and $t_3$ converge to $t_1$. One can show that as long as $F''(t_1) \neq 0$, then this limit $\mathbf{C}$ exists. Basically, $\mathbf{C}$ is determined by its center which is the limit of the centers of the circles $\mathbf{C}(t_1,t_2,t_3)$. The circle $\mathbf{C}$, called the *osculating circle* for the curve at $F(t_1)$, is the best matching circle we are seeking. Its center is called the *center of curvature* and its radius is called the *radius of curvature* of the curve at $F(t_1)$.

**Second geometric definition of curvature:** Again consider the point $F(t_1)$ and let $F(t_2)$ be a nearby point. The angle between the tangent vectors $F'(t_1)$ and $F'(t_2)$ can be used as a measure of how curved the curve is at $F(t_1)$ when $t_2$ is close to $t_1$. See Figure 9.4. More precisely, define the curvature of the curve at $F(t_1)$ to be

$$\lim_{t_2 \to t_1} \frac{\angle(F'(t_1), F'(t_2))}{|F(t_1)F(t_2)|}. \tag{9.4}$$

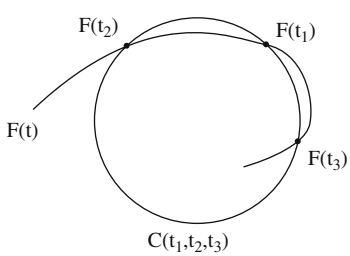



**Figure 9.3.** Curvature in terms of the best approximating circle.

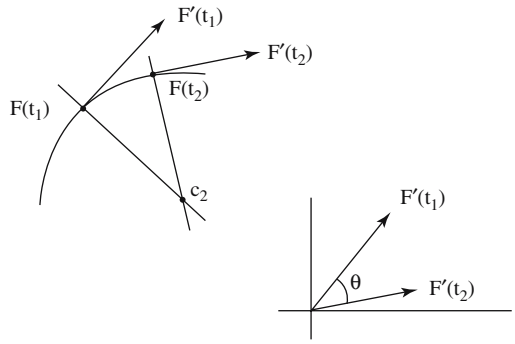



**Figure 9.4.** Curvature in terms of rotating tangent vector.

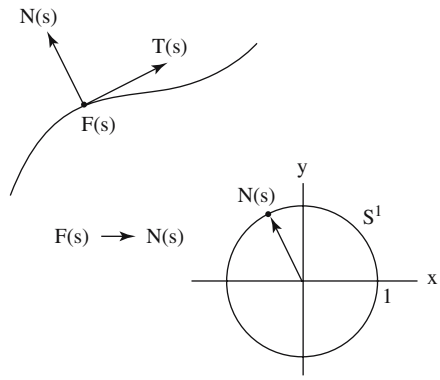



**Figure 9.5.** The Gauss map for planar curves.

Let $\mathbf{c}_2$ be the point that is the intersection of the normal lines to the curve at $F(t_1)$ and $F(t_2)$. One can show that the points $\mathbf{c}_2$ converge to the center of curvature of the curve at $F(t_1)$ and that the expression (9.4) is the reciprocal of the radius of curvature.

In the next part of the discussion it is convenient to switch to arc-length parameterization. Let F(s) be the arc-length parameterization of our curve and let T(s) = F′(s). There are precisely two **unit** vectors at F(s) that are normal to T(s) at F(s). Let N(s) denote the one such that (T(s),N(s)) induces the standard orientation of $\mathbf{R}^2$. This defines N(s) uniquely. In fact, since T(s) is a unit vector, if $T(s) = (T_1(s),T_2(s))$, then $N(s) = (-T_2(s),T_1(s))$.

Now both T(s) and N(s) can be thought of as maps that map the point F(s) on the curve into the unit circle $\mathbf{S}^1$. Thought of in this way, the map N(s) is a special case of what is called the *Gauss map* whose generalization to surfaces plays a fundamental role in the study of surfaces. See Figure 9.5. In our second geometric definition of curvature we could have replace the angle between tangent vectors by the angle between the corresponding normal vectors since they are the same. The Gauss map shows that circles naturally come into the picture when studying curvature. Basically, N(s) relates changes of angles on the curve with the corresponding changes for the mapped curve in the circle.

With this intuitive introduction to curvature we are ready to give a rigorous definition. The definition is surprisingly quite simple and determining the curvature of

a curve is a straightforward computation. It is based on the idea that the rate of change of the tangent is also a measure of curvature.

Recall Proposition 9.2.6, which showed that $T'(s)$ is orthogonal to $T(s)$. It follows that $N(s)$ and $T'(s)$ are multiples of each other.

**Definition.** Given an arc-length parameterized curve $F(s)$ with tangent vector function $T(s) = F'(s)$, the *curvature vector* $K(s)$, the *signed curvature* $\kappa_S(s)$, and the *curvature* $\kappa(s)$ of $F$ at $s$ are defined by

$$K(s) = T'(s) = \kappa_S(s)N(s) \quad \text{and} \quad \kappa(s) = |\kappa_S(s)|.$$

The vector $N(s)$ is called the *principal normal* of $F$ at $s$.

Note that $\kappa(s) = |T'(s)|$, which means that it is easy to compute the size of the curvature.

**9.3.1. Example.** It is easy to check that

$$G(s) = \mathbf{p} + s\frac{\mathbf{pq}}{|\mathbf{pq}|}, \quad s \in [0, |\mathbf{pq}|],$$

is the arc-length parameterization of the segment $[\mathbf{p},\mathbf{q}]$. It follows that the curvature of a straight line is zero since $G''(s) = \mathbf{0}$.

**9.3.2. Example.** Consider the circle of radius $r$ with center the origin for which the function

$$G(s) = \left(r\cos\frac{s}{r}, r\sin\frac{s}{r}\right), \quad s \in [0, 2\pi],$$

is easily checked to be an arc-length parameterization. Clearly,

$$\begin{aligned} T(s) &= G'(s) = \left(-\sin\frac{s}{r}, \cos\frac{s}{r}\right), \\ T'(s) &= \left(-\frac{1}{r}\cos\frac{s}{r}, -\frac{1}{r}\sin\frac{s}{r}\right), \quad \text{and} \\ N(s) &= (-\cos s, -\sin s). \end{aligned}$$

Therefore, $\kappa_S(s) = 1/r$, so that our definition also gives the correct answer for circles.

A more direct way to compute the signed curvature without determining the normal vectors as was done in the definition is:

**9.3.3. Proposition.** With the notation as above

$$\kappa_S(s) = \det\begin{pmatrix} T(s) \\ T'(s) \end{pmatrix}.$$

**Proof.** We can think of $\kappa(s)$, which is the length of the vector $T'(s)$, as the area of the rectangle spanned by the orthogonal vectors $T(s)$ and $T'(s)$, but such an area can be computed from the cross product of the two vectors. Since the z-component of the vectors is 0, the formula works out to the determinant as indicated. It is easy to check that the signs are correct also.

**Note 1.** Our parameterizations will typically be assumed to be **regular**. Recall that "regular" means that the derivative does not vanish and hence avoids various degenerate cases that a zero derivative would cause, such as it being in the denominator of a formula as in the case of curvature.

**Note 2.** Because the arc-length parameterization $G(s)$ of a path is unique given a traversal direction, the values of functions defined in terms of it should really be thought of as associated to the corresponding points on the **path**. Therefore, and this is another reason for using regular parameterizations, if $F(t)$ is a regular parameterization for the same curve, then we can think of such values as functions of t because the change of parameter function between t and s is a one-to-one correspondence. This means that we may use expressions such as "the principal normal at t" or "the curvature at t" and write $N(t)$ or $\kappa(t)$, respectively. We must be a little careful though because, if signed quantities are involved, such as the signed curvature, then the parameter t and the arc-length parameter s need to traverse the curve in the same direction.

Following up on Note 2, since curves are hardly ever presented via their arc-length parameterization because that usually involves complicated formulas, functions defined in terms of arc-length parameterization are not very useful from a computational point of view. It important therefore that one can compute them with respect to arbitrary parameterizations.

To begin with, consider a regular parameterization $F(t)$ for a curve and let $G(s) = F(\phi(s))$ be its arc-length parameterization, where $t = \phi(s)$ and $\phi'(s) > 0$. Now the equation

$$T(s) = G'(s) = F'(\phi(s))\phi'(s) \tag{9.5}$$

together with the facts that $T(s)$ is a unit vector and $\phi'(s) > 0$ imply that

$$\phi'(s) = \frac{1}{|F'(\phi(s))|}. \tag{9.6}$$

Next, let $S(t) = T(\phi^{-1}(t))$. Then $S'(t) = T'(\phi^{-1}(t))\phi^{-1\prime}(t)$, implies that

$$T'(s) = \frac{S'(t)}{|F'(t)|}. \tag{9.7}$$

Equations (9.5)–(9.7) lead to the following formulas:

$$T(s) = \frac{F'(\phi(s))}{|F'(\phi(s))|} = \frac{F'(t)}{|F'(t)|} \tag{9.8}$$

$$K(s) = T'(s) = \frac{S'(t)}{|F'(t)|} \left(\text{often written as } \frac{T'(t)}{|F'(t)|}\right). \tag{9.9}$$

We can also compute the signed curvature directly from a parameterization rather than computing it via Proposition 9.3.3.

**9.3.4. Proposition.** If F(t) is any regular planar curve, then

$$\kappa_S(t) = \frac{F_1'(t)F_2''(t) - F_2'(t)F_1''(t)}{\left(F_1'^2(t) + F_2'^2(t)\right)^{3/2}} = \frac{1}{|F'(t)|^3} \det\begin{pmatrix} F'(t) \\ F''(t) \end{pmatrix}.$$

**Proof.** Let

$$n(t) = \frac{1}{\sqrt{F_1'^2(t) + F_2'^2(t)}} \left(-F_2'(t), F_1'(t)\right) \tag{9.10}$$

be the principal normal to the curve at F(t). Straightforward differentiation of equation (9.10) shows that

$$n'(t) = \frac{F_1'(t)F_2''(t) - F_2'(t)F_1''(t)}{\left(F_1'^2(t) + F_2'^2(t)\right)^{3/2}} F'(t). \tag{9.11}$$

But if $s = \alpha(t)$ and $F(t) = G(\alpha(t))$, where G(s) is arc-length parameterization, then $n(t) = N(\alpha(t))$. It follows that $F'(t) = G'(s)\alpha'(t)$ and $n'(t) = N'(s)\alpha'(t)$. Equation (9.11) and the definition of curvature now implies the result.

Using Proposition 9.3.4, we can rewrite equation (9.11) in the following interesting form

$$n'(t) = \kappa_S(t)F'(t). \tag{9.12}$$

This says that the vector that is the rate of change of the unit normal to a parametric curve is parallel to the tangent vector to the curve and also shows that the two are related by the curvature. This is true whether the parameter is arc-length or not.

The next theorem, the fundamental theorem about curves in the plane, states that the signed curvature function describes a curve completely.

**9.3.5. Theorem.** For any function $\kappa_S(s)$ defined on an interval $[s_0, s_1]$, there is a **unique** (up to rigid motion) regular curve F(s) so that $\kappa_S(s)$ is the signed curvature function and s is the arc-length parameterization of F.

**Proof.** The proof amounts to simply writing down a solution, namely, if we define a function θ(s) by

$$\theta(s) = \int \kappa_S(s)ds + \phi, \tag{9.13}$$

then

$$F(s) = \left(\int \cos\theta(s)ds + a, \int \sin\theta(s)ds + b\right). \tag{9.14}$$

This determines F(s) up to a translation (a,b) and rotation through an angle $\phi$. See [DoCa76] or [Gray98].

A corollary of Proposition 9.3.4 and Theorem 9.3.5 is

**9.3.6. Corollary.** The planar curve F(t) traces out a straight line if and only if $F'(t) \times F''(t) = 0$.

The function $\theta(s)$ in the proof of Theorem 9.3.5 is worth another look because it leads to another interpretation of curvature.

**9.3.7. Theorem.** Let F(t) be a regular curve with domain [a,b]. If

$$\frac{F'(t_0)}{|F'(t_0)|} = (\cos\theta_0, \sin\theta_0)$$

for some $t_0 \in (a,b)$ and some fixed angle $\theta_0$, then there is unique differentiable function $\theta(t)$ on [a,b], so that

$$\frac{F'(t)}{|F'(t)|} = (\cos\theta(t), \sin\theta(t))$$

for all $t \in [a,b]$. Furthermore,

$$\frac{d\theta}{dt} = |F'(t)|\kappa_S(t).$$

**Proof.** See [Gray98].

**Definition.** The function $\theta(t)$ in Theorem 9.3.7 is called the *turning angle* of the curve F(t) determined by $\theta_0$.

The function $\theta(s)$ in the proof of Theorem 9.3.5 is just this same turning angle.

**Third geometric definition of curvature:** The signed curvature is the rate of change of the turning angle for an arc-length parameterized curve.

Next, we consider some curvature related properties of planar curves.

**Definition.** A differentiable curve F(t) defined on [a,b] is said to be a *simple closed curve* if

(1) F is one-to-one on (a,b), and
(2) F and all of its derivatives agree at a and b, that is, F(a) = F(b), F′(a) = F′(b), F″(a) = F″(b), etc.

**Definition.** A curve is said to be *convex* if it lies entirely on one side of each tangent line at every point on the curve.

**9.3.8. Theorem.** A simple closed curve is convex if and only if either $\kappa_S(s) \geq 0$ for all s or $\kappa_S(s) \leq 0$ for all s.

**Proof.** See [DoCa76].

One can also show that a curve is convex if and only if it lies in the boundary of its convex hull.

**Definition.** A *vertex* of a plane curve F(s) is a point where $d\kappa_S/ds = 0$, that is, $\kappa_S$ has a local extremum. (Sometimes a vertex is defined to be a point where $\kappa_S$ has a relative minimum or maximum.)

An ellipse that is not a circle has four vertices, namely those points where the axes of the ellipse meet the ellipse. It turns out that no regular closed convex curve has less than four vertices.

**9.3.9. Theorem.** (The Four Vertex Theorem) A simple closed convex curve has at least four vertices.

**Proof.** See [DoCa76] or [Gray98]. For the polygonal analog of this theorem see [Taba00].

The proof of the Four Vertex Theorem actually shows that the curvature function has at least two local maxima and two local minima. Rephrased in those terms, the theorem has a partial converse.

**9.3.10. Theorem.** Let $\kappa:[a,b] \to \mathbf{R}^2$, $a < b$, be any strictly positive function which, along with all of its derivatives, takes on the same value at a and b. If $\kappa$ is either constant or has at least two maxima and two minima, then there is a simple closed curve $F:[a,b] \to \mathbf{R}^2$ whose curvature function is $\kappa$.

**Proof.** See [Gluc71].

Given a simple closed curve in the plane, the Jordan curve theorem implies that it divides the plane into two regions, a bounded and an unbounded part.

**Definition.** The bounded part of a simple closed curve in the plane is called the *interior* of the curve or the *region bounded by the curve*.

An old problem in geometry is the following:

> **The isoperimetric problem:** Of all simple closed curves in the plane with a given length, which one bounds the largest area?

The isoperimetric problem is solved by the next theorem, which says that the obvious answer, namely the circle, is the correct answer.

**9.3.11. Theorem.** (The Isoperimetric Inequality) Let L be the length of a simple closed curve and let A be the area of the region bounded by the curve. Then

$$L^2 - 4\pi A \geq 0,$$

and equality holds if and only if the curve is a circle.

**Proof.** See [DoCa76].

We finish this section with one last geometric definition of curvature.

**Fourth geometric definition of curvature:** If F(s) is the arc-length parameterization for a curve defined on [0,L], then we know that the tangent vectors T(s) have unit length. As indicated earlier, we can think of T as a mapping of [0,L] to the unit circle. Tangent vectors are the more natural choice for what we are trying to do here rather than the well-known Gauss map of normals, but we would accomplish the same thing with either. Given a partition $P = (s_0, s_1, \ldots, s_n)$ of [0,s], let

$$\Theta_i = \angle_s(T(t_i), T(t_{i+1}))$$

denote the signed angle between $T(s_i)$ and $T(s_{i+1})$. Define

$$k_T(s) = \lim_{|P|\to 0} \sum_{i=0}^{n-1} \Theta_i, \tag{9.15}$$

where |P| is the norm of P. See Figure 9.5 again.

**Definition.** If the limit $\kappa_T(s)$ exists, it is called the *total curvature* of the curve F|[0,s]. The value $\kappa_T(L)$ will be called the *total curvature* of F and denoted simply by $\kappa_T$.

The next proposition relates the total curvature function to the signed curvature function.

**9.3.12. Proposition.** $\dfrac{d}{ds}\kappa_T(s) = \kappa_S(s)$.

**Proof.** See [Stok69], [Spiv70b], or [Gray98].

The total curvature of a closed curve is closely related to a topological invariant, namely the degree of an associated map. Given a simple closed curve F(s), then F′(s) = T(s) can be thought of as a map of the circle to itself in a natural way since the interval [0,L] with the endpoints identified is just a circle. As such, we can talk about the degree of F′. See Section 7.5.1 or 8.11.

**9.3.13. Theorem.** If F(s) is the arc-length parameterization of a closed curve of length L, then

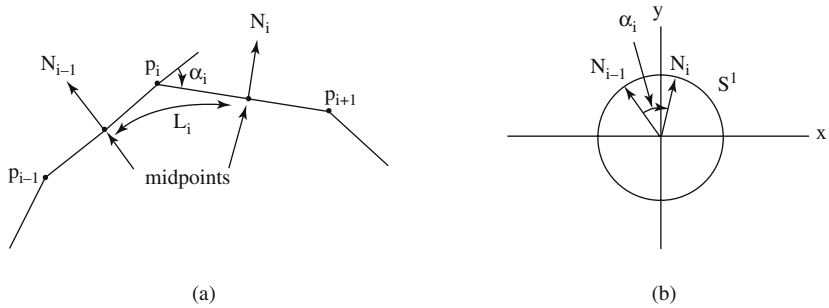



**Figure 9.6.** The curvature at vertices of polygonal curves.

$$\kappa_T(L) = 2\pi \deg T.$$

**Proof.** See [Stok69], [Spiv70b], or [Gray98].

Theorem 9.3.13 can be thought of as a kind of integrality theorem because it states that $\kappa_T/2\pi$ is an integer.

**9.3.14. Corollary.** For a simple closed curve $\kappa_T = 2\pi$.

Closely related to the total curvature is the concept of the winding number of a curve F(t) about a point $\mathbf{p}$. This measures the angle that the vector $F(t) - \mathbf{p}$ sweeps out around the origin. See [AgoM76].

This section described several ways to define the curvature of smooth planar curves. In closing, we would like to indicate how one could define curvature for polygonal curves. In some sense, one can think of the polygonal concept as a precursor of the smooth one. Consider a polygonal curve defined by a sequence of points $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_n$. Figure 9.6(a) shows an example. Any curvature for such a curve would be concentrated at the vertices and zero elsewhere. At a vertex $\mathbf{p}_i$ it is natural to use the signed angle

$$\alpha_i = \angle_s(\mathbf{p}_{i-1}\mathbf{p}_i, \mathbf{p}_i\mathbf{p}_{i+1})$$

between the directed segments as a measure of how much the curve is turning. But the lengths of the segments also play a role, so we need to normalize things. Define the *curvature* $K_i$ at $\mathbf{p}_i$ by

$$K_i = \frac{\alpha_i}{L_i}, \quad \text{where} \quad L_i = \frac{1}{2}(|\mathbf{p}_{i-1}\mathbf{p}_i| + |\mathbf{p}_i\mathbf{p}_{i+1}|).$$

If $\mathbf{N}_i$ is the "outward" unit normal to the segment $[\mathbf{p}_{i-1}, \mathbf{p}_i]$, then $\alpha_i$ is the length of the arc from $\mathbf{N}_{i-1}$ to $\mathbf{N}_i$ in the unit circle $\mathbf{S}^1$. See Figure 9.6(b). This curvature has some interesting properties and is the one-dimensional analog of the Gauss curvature for surfaces defined in Section 9.9. The reader should look back at these comments when

reading that section. What we are doing here is comparing the length of arcs of the curve with the length its Gauss map "traces out" in $\mathbf{S}^1$. If the polygonal curve were an approximation of a smooth one, then one would be able to show that the polygonal curvature approximated the smooth curvature defined earlier. See [Call86].

## 9.4 The Geometry of Space Curves

Next, we consider curves in $\mathbf{R}^3$. Such curves are also called *space curves*. Let

$$F: [0,L] \to \mathbf{R}^3$$

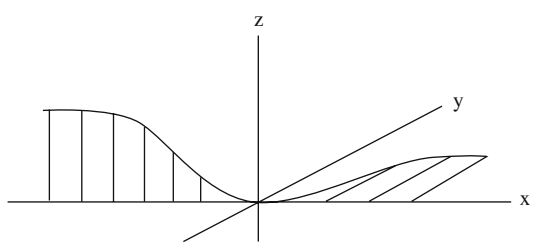
be a parameterization of one of these curves.

**Definition.** If F(s) is the arc-length parameterization and T(s) = F′(s), then the *curvature vector* K(s) and the *curvature* $\boldsymbol{\kappa}(s)$ to the curve F at the point F(s) are defined by

$$K(s) = T'(s) \quad \text{and} \quad \boldsymbol{\kappa}(s) = |K(s)|.$$

Note that there is no definition of a signed curvature for space curves. Space curves are only assigned a nonnegative curvature function.

**A geometric definition of the curvature of a space curve:** The approach is again via best matching circles, but there is more to show now. Given an arbitrary parameterization F(t), define circles $\mathbf{C}(t_1,t_2,t_3)$ through $F(t_i)$ as before. These circles may now lie in different planes. Fortunately, one can show that if $F''(t) \neq 0$, then, as the $t_i$ approach t, the planes determined by the $F(t_i)$ approach the plane generated by F(t) and F″(t) in the limit. Furthermore, the circles $\mathbf{C}(t_1,t_2,t_3)$ approach a limiting circle **C** that lies in this plane. The curvature at F(t) is then the reciprocal of the radius of this circle. The $C^\infty$ parameterization F(t) in Figure 9.7 shows that the hypothesis $F''(t) \neq 0$ is needed because otherwise there might not be any limiting plane or circle.




$$\begin{aligned} F(t) &= (0,0,0) && , t = 0 \\ &= (t,\exp(-1/t^2),0) && , t > 0 \\ &= (t,0,\exp(-1/t^2) && , t < 0 \end{aligned}$$

**Figure 9.7.** Why $F''(t) \neq 0$ is needed for a unique best matching circle.

**Note:** As we just indicated, to get a well-defined geometric definition of curvature for a space curve F(t) we need to assume that F″(t) does not vanish. For that reason and the need for that assumption in other formulas related to geometric properties of curves, the condition $F''(t) \neq 0$ is often assumed implicitly in discussions about curves, just like the condition of regularity. On a related point of terminology, given a space curve F(t), a point $\mathbf{p} = F(t)$ with the property that $F''(t) = 0$ is often referred to as a point at which the curve is *locally flat*. The justification for this is that a curve whose second derivative vanishes in a neighborhood of a point is in fact a straight line in that neighborhood.

**9.4.1. Proposition.** For an arbitrary regular curve F(t) in $\mathbf{R}^3$

$$\boldsymbol{\kappa}(t) = \frac{|F'(t) \times F''(t)|}{|F'(t)|^3}.$$

**Proof.** See [Spiv70b]. Compare this formula to the one for plane curves in Proposition 9.3.4.

**Definition.** A point of a curve where the curvature vanishes is called an *inflection point*.

**Definition.** If F(s) is an arc-length parameterization and if $T(s) = F'(s)$, then the *principal normal* of F at s, denoted by N(s), and the *binormal* of F at s, denoted by B(s), are defined by

$$T'(s) = \boldsymbol{\kappa}(s)N(s) \quad \text{and} \quad B(s) = T(s) \times N(s),$$

where $\boldsymbol{\kappa}(s)$ is the curvature.

Clearly, both N(s) and B(s) are unit vectors and the orthonormal basis (T(s),N(s),B(s)) determines the standard orientation of $\mathbf{R}^3$.

**Definition.** The tuple (T(s),N(s),B(s)) is called the *Frenet frame* or *moving trihedron* to the curve F(s) at s or the point F(s). The Frenet frame or moving trihedron at a point of an arbitrary regular parameterization is the Frenet frame or moving trihedron at that point of the induced arc-length parameterization.

**Definition.** The *osculating plane* of F at s is the plane at F(s) generated by T(s) and N(s). The *normal plane* of F at s is the plane at F(s) generated by N(s) and B(s). The *rectifying plane* of F at s is the plane at F(s) generated by T(s) and B(s).

**9.4.2. Proposition.** $B'(s) = -\boldsymbol{\tau}(s)N(s)$ for some function $\boldsymbol{\tau}(s)$.

**Proof.** First, $B \bullet B = 1$ implies that

$$B' \bullet B = 0.$$

Second, $B \bullet T = 0$ implies that

$$B' \bullet T = -B \bullet T' = -B \bullet \kappa N = 0.$$

Therefore, B′ is orthogonal to both B and N. Since (T,N,B) is an orthonormal basis, it follows that B′ must be some multiple of N.

**Definition.** The function $\tau(s)$ in Proposition 9.4.2 is called the *torsion* of F at s.

The minus sign in Proposition 9.4.2 is there so that (T(s),N(s),F‴(s)) determines the standard orientation whenever $\tau(s) > 0$. See equation 9.16 below.

**9.4.3. Example.** The curve in $\mathbf{R}^3$ parameterized by

$$F(t) = (a\cos t, a\sin t, bt),$$

where a and b are nonzero constants is called a *helix*. Let us compute the curvature and torsion of this helix.

**Solution.** First, we find the arc-length parameterization G(s) of the helix using the method described in Section 9.2. We get

$$G(s) = F\left(\frac{s}{D}\right) = \left(a\cos\frac{s}{D}, a\sin\frac{s}{D}, b\frac{s}{D}\right), \quad \text{where} \quad D = \sqrt{a^2+b^2}.$$

It follows that

$$T(s) = \left(-\frac{a}{D}\sin\frac{s}{D}, \frac{a}{D}\cos\frac{s}{D}, \frac{b}{D}\right), \quad T'(s) = \left(-\frac{a}{D^2}\cos\frac{s}{D}, -\frac{a}{D^2}\sin\frac{s}{D}, 0\right),$$

$$N(s) = \left(-\cos\frac{s}{D}, -\sin\frac{s}{D}, 0\right), \quad \text{and} \quad B(s) = \frac{a}{|a|}\left(\frac{b}{D}\sin\frac{s}{D}, -\frac{b}{D}\cos\frac{s}{D}, \frac{a}{D}\right).$$

It is now easy to show that

$$\kappa(s) = \frac{|a|}{D^2} \quad \text{and} \quad \tau(s) = \frac{b}{D^2}.$$

and we are done.

Since B is orthogonal to the osculating plane, and $|\tau(s)| = |B'(s)|$, we can think of $\tau(s)$ as measuring the rate at which the osculating plane is changing. In other words, $\tau(s)$ measures by how much the curve deviates from being planar.

**9.4.4. Theorem.** $\tau(s) = 0$ if and only if the curve $F(s)$ is a planar curve.

**Proof.** See [Lips69].

**9.4.5. Proposition.** If the curve $F(s)$ is an arc-length parameterization, then

$$\tau(s) = \frac{1}{\kappa^2(s)}(F'(s) \times F''(s)) \bullet F'''(s) = \frac{1}{\kappa^2(s)} \det \begin{pmatrix} F'(s) \\ F''(s) \\ F'''(s) \end{pmatrix}. \tag{9.16}$$

**Proof.** This is a straightforward computation using the various definitions and vector identities, namely,

$$\begin{aligned}
\tau &= -B' \bullet N \\
&= -(T \times N)' \bullet N = -[T \times N' + T' \times N] \bullet N = -(T \times N') \bullet N \\
&= -\left(T \times \left(\frac{T'}{\kappa}\right)'\right) \bullet \left(\frac{T'}{\kappa}\right) = -\left(T \times \frac{\kappa T'' - T'\kappa'}{\kappa^2}\right) \bullet \left(\frac{T'}{\kappa}\right) = -\left(T \times \frac{T''}{\kappa}\right) \bullet \left(\frac{T'}{\kappa}\right) \\
&= -\frac{1}{\kappa^2}(T \times T'') \bullet T' = -\frac{1}{\kappa^2}(F' \times F'') \bullet F'''.
\end{aligned}$$

Next, we want to determine the well-known equations that relate the vector fields $T'$, $N'$, and $B'$ to the vector fields $T$, $N$, and $B$ along the curve. We already know that $T' = \kappa N$ and $B' = -\tau N$. Let

$$N' = aT + bN + cB.$$

Since $N \bullet N = 1$, it follows that $N' \bullet N = 0$, that is, $N'$ is orthogonal to $N$. This means that $b = 0$. But $N \bullet T = 0$ implies that

$$a = N' \bullet T = -N \bullet T' = -N \bullet \kappa N = -\kappa,$$

and $N \bullet B = 0$ implies that

$$c = N' \bullet B = -N \bullet B' = -N \bullet (-\tau N) = \tau.$$

Collecting these facts gives us the well-known theorem below due to Serret (1851) and Frenet (1847):

**9.4.6. Theorem.** (The Serret-Frenet Formulas) The following equations hold for arc-length parameterizations:

$$\begin{aligned}
T' &= \phantom{-\kappa T} \quad \kappa N \\
N' &= -\kappa T \phantom{\kappa N} \quad + \tau B \\
B' &= \phantom{-\kappa T} \quad -\tau N
\end{aligned}$$

The Serret-Frenet formulas are the key to proving the fundamental theorem for space curves:

**9.4.7. Theorem.** Let $\boldsymbol{\kappa}, \boldsymbol{\tau}:[0,L] \to \mathbf{R}$ be continuous functions with $\boldsymbol{\kappa}(s) > 0$. Then there is a **unique** (up to a rigid motion) curve $F:[0,L] \to \mathbf{R}^3$ parameterized by arc-length whose curvature and torsion functions are the functions $\boldsymbol{\kappa}$ and $\boldsymbol{\tau}$, respectively.

**Proof.** The existence part of this theorem involves solving differential equations associated to the Serret-Frenet formulas. The uniqueness part is easier. See [Lips69] or [Spiv70b].

Since one is rarely given arc-length parameterizations, it is convenient to have formulas for the curvature and torsion of space curves with respect to other parameterizations.

Let $F(t)$ be an arbitrary regular curve and assume that $F(t) = G(\alpha(t))$, where $G(s)$ is the arc-length parameterization and $s = \alpha(t)$. Just like in the planar case, we shall assume that the basic geometric properties defined by $G(s)$, such as curvature and torsion, can be assumed to be associated to **points** on the curve and hence to $F(t)$. If $\boldsymbol{\kappa}_G(s)$ and $\boldsymbol{\tau}_G(s)$ and the curvature and torsion functions associated to $G(s)$ and if $(T_G(s),N_G(s),B_G(s))$ is the Frenet frame for $G(s)$, then define the corresponding functions for $F(t)$ by

$$\begin{aligned}
\boldsymbol{\kappa}(t) &= \boldsymbol{\kappa}_G(\alpha(t)),\\
\boldsymbol{\tau}(t) &= \boldsymbol{\tau}_G(\alpha(t)),\\
T(t) &= T_G(\alpha(t)),\\
N(t) &= N_G(\alpha(t)), \quad \text{and}\\
B(t) &= B_G(\alpha(t)).
\end{aligned} \tag{9.17}$$

**9.4.8. Theorem.** (The generalized Serret-Frenet Formulas) Given a regular curve $F(t)$, then the functions defined by Equations (9.17) satisfy

$$\begin{aligned}
T' &= \phantom{-v\boldsymbol{\kappa}T} \quad v\boldsymbol{\kappa}N\\
N' &= -v\boldsymbol{\kappa}T \quad\quad + v\boldsymbol{\tau}B\\
B' &= \phantom{-v\boldsymbol{\kappa}T} \quad -v\boldsymbol{\tau}N
\end{aligned}$$

where $v(t) = |F'(t)|$ is the speed function of $F(t)$.

**Proof.** This follows from the chain rule applied to the functions in equations (9.17) and Theorem 9.4.6.

**9.4.9. Theorem.** If $F(t)$ is an arbitrary regular curve with $F''(t) \neq 0$ (equivalently, nonzero curvature), then the functions defined by Equations (9.17) satisfy the following identities:

(1) $T(t) = \dfrac{F'(t)}{|F'(t)|}$

(2) $B(t) = \dfrac{F'(t) \times F''(t)}{|F'(t) \times F''(t)|}$

(3) $N(t) = B(t) \times T(t)$

(4) $\kappa(t) = \dfrac{|F'(t) \times F''(t)|}{|F'(t)|^3}$

(5) $\tau(t) = \dfrac{(F'(t) \times F''(t)) \bullet F'''(t)}{|F'(t) \times F''(t)|^2} = \dfrac{1}{|F'(t) \times F''(t)|^2} \det \begin{pmatrix} F'(t) \\ F''(t) \\ F'''(t) \end{pmatrix}.$

**Proof.** This is easy if one copies what we did for arc-length parameterizations above. See [Gray98] or [Spiv70b].

An immediate corollary of Theorems 9.4.4 and 9.4.9 is

**9.4.10. Theorem.** The regular curve F(t) is a planar curve if and only if

$$(F'(t) \times F''(t)) \bullet F'''(t) = \det \begin{pmatrix} F'(t) \\ F''(t) \\ F'''(t) \end{pmatrix} = 0.$$

Finally, we note that the Serret-Frenet formulas generalize to curves in $\mathbf{R}^n$.

**9.4.11. Theorem.** Let F(s) be a curve in $\mathbf{R}^n$, $n \geq 3$, parameterized by arc-length. If the vectors $F'(s), F''(s), \ldots, F^{(n)}(s)$ are linearly independent, then at each point F(s) on the curve there is an orthonormal basis of vectors $u_1(s), u_2(s), \ldots, u_n(s)$ with the function $u_i(s)$ satisfying the differential equations

$$\begin{aligned}
u_1' &= \qquad\qquad \kappa_1 u_2 \\
u_2' &= -\kappa_1 u_1 + \kappa_2 u_3 \\
u_3' &= -\kappa_2 u_2 + \kappa_3 u_4 \\
\cdot \quad & \quad \cdot \qquad \cdot \\
\cdot \quad & \quad \cdot \qquad \cdot \\
u_{n-1}' &= -\kappa_{n-2} u_{n-2} + \kappa_{n-1} u_n \\
u_n' &= -\kappa_{n-1} u_{n-1}
\end{aligned}$$

**Proof.** See [Spiv75].

**Definition.** The functions $u_1, u_2, \ldots, u_n$ in Theorem 9.4.11 are called a *Frenet basis* and the functions $\kappa_i$ are called *generalized curvatures* for the curve F(s).

If n = 3, then $\kappa_1$ is the ordinary curvature of a curve in 3-space and $\kappa_2$ is the torsion.

## 9.5 Envelopes of Curves

Envelopes of families of curves or surfaces were an important part of classical differential geometry, especially in the development of the concept of a connection which is fundamental to modern differential geometry. See [Spiv75]. They also appear in the context of cyclides and developable surfaces. Even so, they have only been studied to a limited degree, mainly envelopes of circles, planes, and spheres. An analysis can get very tricky even in very simple sounding situations. We consider envelopes here because of their relevance to CAGD.

This section looks at envelopes of planar curves. Section 9.11 considers envelopes of surfaces.

**Definition.** Let $\alpha_t : [0,1] \to \mathbf{R}^2$ be a one-parameter family of curves in the plane defined by $\alpha_t(u) = \alpha(u,t)$ for some $C^\infty$ function $\alpha : [0,1] \times [0,1] \to \mathbf{R}^2$. An *envelope* of this family is defined to be a curve $p(u)$ that is not a member of this family but that is tangent to some member of the family at every point.

See Figure 9.8(a). Unfortunately, the envelope can have bad singularities like cusps even if $\alpha$ is a nice function. See Figure 9.8(b), which shows the envelope of normals to an ellipse whose ends are the centers of the osculating circles.

The classical approach to studying the envelope $p(u)$ is geometric. Basically, one defines $p(u)$ to be the limit as $\varepsilon$ approaches 0 of the intersections of $\alpha_t(u)$ and $\alpha_t(u + \varepsilon)$. Such a definition can have serious problems in general, but seems to work in many cases of interest.

Let us start off with a special case that will not only make the argument clearer but will also be used for the general case since that case will reduce to this one. Assume that the curves $\alpha_t$ are graphs of functions $f_t$, that is,

$$\alpha(u,t) = (u,v), \quad \text{where } v = f(u,t) = f_t(u).$$

If the curve $\alpha_t$ intersects the curve $\alpha_{t+h}$ at the point $(u_h, f(u_h,t)) = (u_h, f(u_h, t + h))$, then

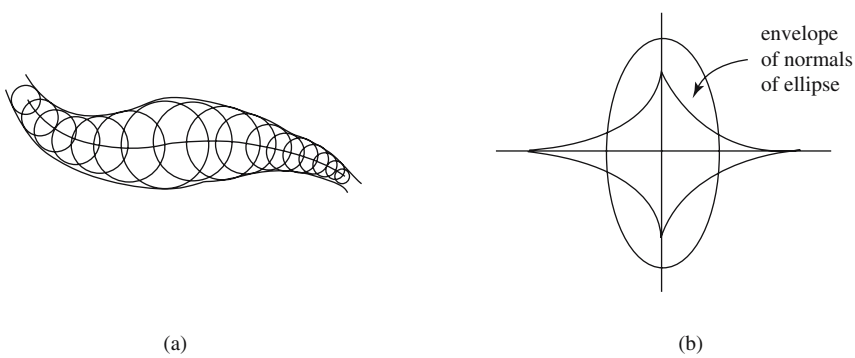



**Figure 9.8.** Envelopes.

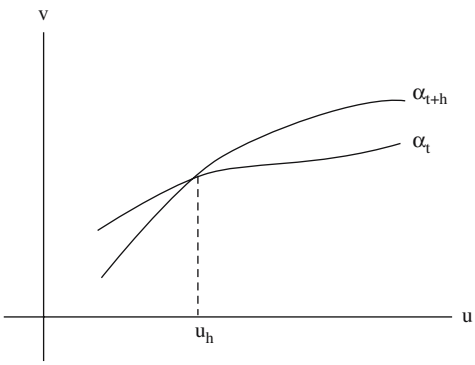



**Figure 9.9.** Intersecting envelopes.

$$0 = \frac{f(u_h, t+h) - f(u_h, t)}{h}.$$

See Figure 9.9. If we assume that the numbers $u_h$ approach a limit u(t) as h approaches 0, then we must have

$$D_2 f(u(t), t) = \frac{\partial f}{\partial t}(u(t), t) = 0. \tag{9.18}$$

The envelope p(t) will be defined by

$$p(t) = (u(t), f(u, u(t))).$$

Next, consider the general case. The functions $\alpha_t$ may not be the graphs of functions of u, but, locally, we can find a coordinate system so that they will be graphs of functions with respect to one of the new coordinate variables (simply rotate the standard coordinate system appropriately). Therefore, if

$$\alpha(u, t) = (\alpha_1(u, t), \alpha_2(u, t)),$$

we can find a function T(u,t) so that

$$\alpha_1(T(u, t), t) = u. \tag{9.19}$$

Define

$$f(u, t) = \alpha_2(T(u, t), t).$$

Note that the function $\beta(u,t) = \alpha(T(u,t),t)$ is then the graph of the function f because

$$\beta(u, t) = \alpha(T(u, t), t) = (\alpha_1(T(u, t), t), \alpha_2(T(u, t), t)) = (u, f(u, t)).$$

Applying equation (9.18) to f(u,t) gives

$$0 = D_1\alpha_2(T(u,t),t)D_2T(u,t) + D_2\alpha_2(T(u,t),t). \tag{9.20}$$

On the other hand, differentiating equation (9.19) gives

$$D_1\alpha_1(T(u,t),t)D_2T(u,t) + D_2\alpha_1(T(u,t),t) = 0,$$

that is,

$$D_2T(u,t) = -\frac{D_2\alpha_1(T(u,t),t)}{D_1\alpha_1(T(u,t),t)}.$$

After substituting into equation (9.20), we finally obtain

$$(D_1\alpha_2\, D_2\alpha_1 - D_1\alpha_1\, D_2\alpha_2)(T(u,t),t) = 0.$$

We conclude that the envelope in general should consist of the points $\alpha(u,t)$ where $(u,t)$ satisfies

$$\det\,(D_i\alpha_j(u,t)) = 0. \tag{9.21}$$

Getting criterion (9.21) involved somewhat loose reasoning, but independent of how we got equation (9.21), one can show that the envelope we want **must** be a subset of the points defined by that equation.

**9.5.1. Theorem.** Equation (9.21) is a necessary condition that points of an envelope defined by a family of curves $\alpha(u,t)$ must satisfy. Alternatively, the condition can be expressed as

$$\frac{\partial\alpha}{\partial u} \times \frac{\partial\alpha}{\partial t} = 0. \tag{9.22}$$

**Proof.** See [Spiv75] or [Stok69].

**Note:** Condition (9.21) is only a **necessary** condition for an envelope but not a **sufficient** condition! See [Stok69].

**9.5.2. Example.** To determine the envelope of circles of radius 1 centered on the line $y = x$ in the plane. See Figure 9.10.

**Solution.** If we parameterize the circles by angles, then we can apply Theorem 9.5.1 to the map $\alpha(u,t)$ defined by

$$\alpha(u,t) = (t,t) + (\cos u, \sin u).$$

Since

$$D_1\alpha_2 = \cos u, \quad D_2\alpha_1 = 1, \quad D_1\alpha_1 = -\sin u, \quad \text{and} \quad D_2\alpha_2 = 1,$$

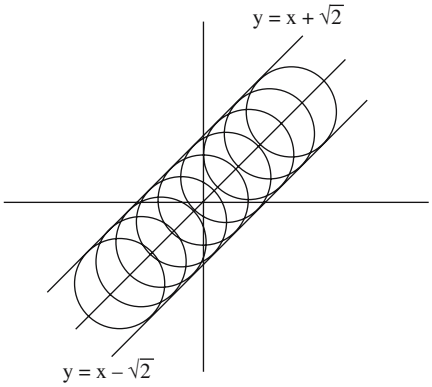



**Figure 9.10.** An envelope of circles.

Equation (9.21) implies the constraint

$$\cos u + \sin u = 0,$$

that is, $u = -\pi/4$ or $3\pi/4$. The envelope therefore consists of a subset of the points

$$\left\{(t,t)+\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right) \mid t \in \mathbf{R}\right\} \cup \left\{(t,t)+\left(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right) \mid t \in \mathbf{R}\right\},$$

which corresponds to the lines $y = x + \sqrt{2}$ and $y = x - \sqrt{2}$. This is clearly a correct answer.

Finally, we look at the problem of finding the envelope of a family of curves in the xy-plane defined implicitly in the form

$$\alpha(x,y,t) = 0. \tag{9.23}$$

Think of Equation (9.23) as defining an equation in x and y for each fixed t.

**9.5.3. Theorem.** If a family of curves is defined implicitly by equation (9.23), then the envelope of that family is a subset of the set of points (x,y) satisfying (9.23) and

$$\frac{\partial \alpha}{\partial t}(x,y,t) = 0. \tag{9.24}$$

**Proof.** One can give an argument similar to the one for Theorem 9.5.1 by thinking of y in equation (9.23) as a function of x and t (possibly after some change in coordinates). Alternatively, one can look at the intersection of two of the curves f(x,y,t) and f(x,y,t + h) and let h go to zero. See also [Brec92].

**9.5.4. Example.** We redo Example 9.5.2 by expressing the family of circles implicitly as

$$\alpha(x,y,t) = (x-t)^2 + (y-t)^2 - 1 = 0. \tag{9.25}$$

**Solution.** In this case, equation (9.24) reduces to

$$x + y = 2t. \tag{9.26}$$

Eliminating t from (9.25) and (9.26) leads to

$$(x-y)^2 = 2,$$

which gives the same answer as in Example 9.5.2.

## 9.6 Involutes and Evolutes of Curves

**Definition.** Given a curve **C**, the union of all tangent lines of **C** is called the *tangent surface* of **C**. An *involute* of **C** is any curve that lies on the tangent surface of **C** and that intersects all the tangent lines orthogonally.

See Figure 9.11. Let p(s) be the arc-length parameterization of a curve **C** and (T(s),N(s),B(s)) its moving trihedron. Let **C*** be an involute of **C**. By definition, there is a function $\alpha(s)$, so that

$$q(s) = p(s) + \alpha(s)T(s)$$

is a parameterization of **C***. Furthermore, $q'(s)$ is orthogonal to T(s), that is,

$$0 = q'(s) \bullet T(s) = (T(s) + \alpha'(s)T(s) + \alpha(s)T'(s)) \bullet T(s) = 1 + \alpha'(s).$$

This implies that $\alpha(s) = -s + c$, for some constant c, proving

**9.6.1. Theorem.** The involutes of a curve p(s) are precisely those curves that admit a parameterization of the form

$$q(s) = p(s) + (c-s)T(s), \tag{9.27}$$

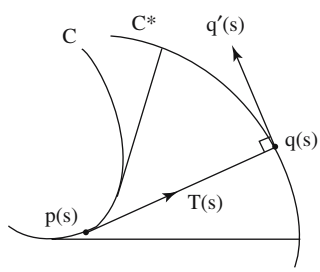



**Figure 9.11.** An involute.

where c is a constant.

The parameterization q(s) in Equation (9.27) is not regular at a point of inflection of p(s) because the derivative

$$q'(s) = T(s) - T(s) + (c - s)T'(s) = (c - s)\boldsymbol{\kappa}(s)N(s)$$

vanishes when $\boldsymbol{\kappa}(s) = 0$.

**9.6.2. Theorem.** The curvature $\boldsymbol{\kappa}^*$ of the involute $\mathbf{C}^*$ is given by

$$\boldsymbol{\kappa}^* = \frac{\boldsymbol{\kappa}^2 + \boldsymbol{\tau}^2}{(c - s)^2 \boldsymbol{\kappa}^2}.$$

**Proof.** See [Lips69].

There is a nice geometric description of the involute. Assume that p(s) is defined on the interval [0,c] and that we have wrapped a string of length c along the curve **C** starting at p(0) and ending at p(c). If we were to unwrap the string holding it taut, then the end of the string generates the involute defined in equation (9.27). See Figure 9.12(a). Another fact that one can easily show is that the distance between two involutes is constant. See Figure 9.12(b).

**Definition.** If a curve $\mathbf{C}^*$ is the involute of a curve **C**, then **C** is called an *evolute* of $\mathbf{C}^*$.

Again, let p(s) be the arc-length parameterization of a curve **C** and (T(s),N(s),B(s)) its moving trihedron. Let $\mathbf{C}^*$ be an evolute of **C**. There exist functions $\alpha(s)$ and $\beta(s)$, so that

$$q(s) = p(s) + \alpha(s)N(s) + \beta(s)B(s) \tag{9.28}$$

is a parameterization of $\mathbf{C}^*$. The reason that there is no T(s) term is that, by definition, q(s) – p(s) is orthogonal to T(s). See Figure 9.13.

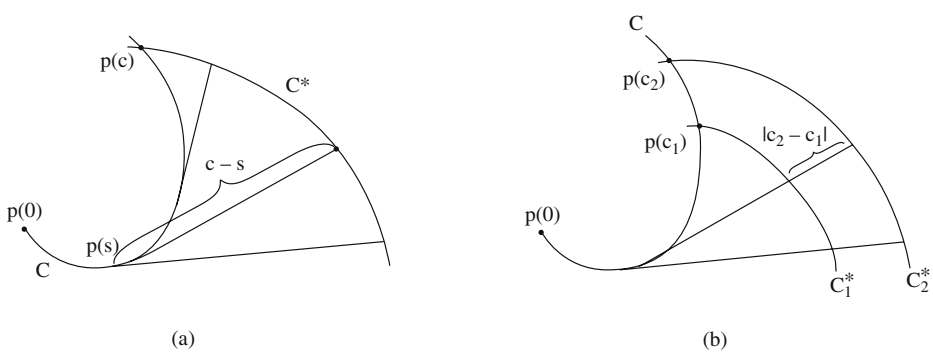
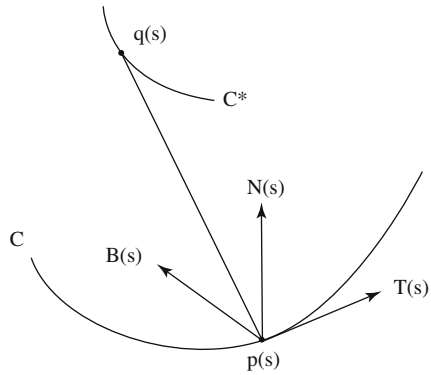



**Figure 9.12.** Generating involutes by unwrapping a string.

**Figure 9.13.** Parameterizing an evolute.




**9.6.3. Theorem.** The evolutes of a curve p(s) are precisely those curves that admit a parameterization of the form

$$q(s) = p(s) + \frac{1}{\kappa(s)}N(s) + \frac{1}{\kappa(s)}\cot\left(\int_0^s \tau + c\right)B(s) \tag{9.29}$$

for some constant c.

**Proof.** Differentiate equation (9.28) and use the fact that q′(s), being tangent to **C***, is a multiple of q(s) – p(s). See [Lips69].

**9.6.4. Corollary.** The evolutes of a planar curve p(s) are precisely those curves that admit a parameterization of the form

$$q(s) = p(s) + \frac{1}{\kappa(s)}N(s) + \frac{c}{\kappa(s)}B(s)$$

for some constant c. The parameterization will be regular if $\kappa'(s)$ is nonzero.

The case of an evolute of a planar curve for which the constant c is zero is especially interesting. That evolute actually lies in the same plane as the curve and is the **only** evolute of the curve in that plane.

**Definition.** Assume that the planar curve p(s) has nonzero curvature. The evolute

$$q(s) = p(s) + \frac{1}{\kappa(s)}N(s) \tag{9.30}$$

is called the *plane evolute* of p(s).

The plane evolute is the locus of centers of curvature of p(t). Because it is unique, the adjective "plane" is often dropped and one talks about "the" evolute of the curve in that plane.

## 9.7 Parallel Curves

Parallel curves and surfaces are topics that, if discussed at all, are usually only mentioned briefly or are relegated to the exercises in books on differential geometry. They are however important for CAGD where they are called offset curves and surfaces and for that reason we shall devote more time to them here. This section will summarize some of the relevant properties of planar parallel curves. Section 9.14 will consider parallel surfaces.

Consider a planar regular parametric curve

$$p(t) = (x(t), y(t)),$$

and let n(t) be the principal normal to the curve at p(t).

**Definition.** A *parallel curve* to p(t) is a parameterized curve of the form

$$p_d(t) = p(t) + dn(t), \tag{9.31}$$

where d is any real number. In CAGD a parallel curve is called an *offset curve*.

Figure 9.14 shows some parallel curves. The figures also show some potential problems. Parallel curves could have self-intersections even if the original does not. They may intersect the original curve.

The tangent vector to $p_d(t)$ is parallel to the tangent vector to p(t). In fact,

$$p_d'(t) = p'(t)(1 + \kappa(t)d), \tag{9.32}$$

where $\kappa(t)$ is the curvature function of p(t). This follows from equation (9.11). It is easy to see from Proposition 9.3.3 and equation (9.32) that the curvature function $\kappa_d(t)$ for $p_d(t)$ is defined by

$$\kappa_d(t) = \frac{\kappa(t)}{1 + \kappa(t)d}. \tag{9.33}$$


**Figure 9.14.** Parallel curves.

Note that even if the original curve p(t) is a regular curve, the parallel curve $p_d(t)$ may not be. Therefore because of their importance in CAGD we shall analyze their singularities and other special points. First of all, there may be cusps at those points where

$$1 + \boldsymbol{\kappa}(t)d = 0.$$

**Definition.** A point on the parallel curve $p_d(t)$, $d \neq 0$, is called a (ordinary) *cusp* if

$$\boldsymbol{\kappa}(t) = -\frac{1}{d}$$

and $\boldsymbol{\kappa}'(t) \neq 0$. A parallel curve without cusps is called *nondegenerate*.

**Definition.** A point on the parallel curve $p_d(t)$ is called an *extraordinary cusp* if

$$\boldsymbol{\kappa}(t) = -\frac{1}{d}$$

**and** $\boldsymbol{\kappa}'(t) = 0$ and $\boldsymbol{\kappa}''(t) \neq 0$.

One can show that the curvature of the parallel curve goes to infinity as we approach an extraordinary cusp.

**9.7.1. Theorem.** The cusps of the parallel curve $p_d(t)$, $d \neq 0$, lie on the evolute of p(t) and meet that curve orthogonally. The extraordinary cusps of the parallel curve coincide with the cusps of the evolute.

**Proof.** See Figure 9.15. The first part is clear from the definition of the plane evolute. See [FarN90a] for a proof of the second.

Now cusps and extraordinary cusps of a parallel curve correspond to differentiability discontinuities that did not exist on the original curve. There are other special points that do not correspond to discontinuities. One of these is an inflection point. This is where the curvature vanishes. We can find such points for a parallel curve from Equation (9.33). Another special point is a vertex of a curve. Recall that a vertex on a curve is a place where the derivative of the curvature function vanishes. Therefore,

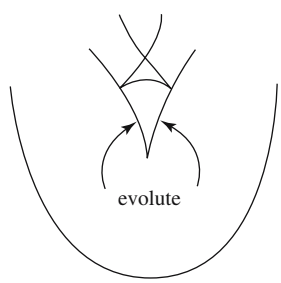



**Figure 9.15.** Cusps of parallel curves lie on evolute.

to find the vertices of a parallel curve we need a formula for the derivative of its curvature function. Use equations (9.3) and (9.33) to get

$$\frac{d\kappa_d}{ds} = \frac{d\kappa_d}{dt}\frac{dt}{ds} = \frac{1}{|p_d'(t)|}\frac{d}{dt}\left(\frac{\kappa}{1+\kappa d}\right).$$

This and (9.32) leads to the equation

$$\frac{d\kappa_d}{ds_d} = \frac{1}{(1+\kappa d)^3}\frac{d\kappa}{ds}. \tag{9.34}$$

For the final special point, let p(t) = (x(t),y(t)) be a regular curve in the plane.

**Definition.** A *turning point* for p(t) is a point where p′(t) is either vertical or horizontal, that is, either x′(t) = 0, y′(t) ≠ 0, or x′(t) ≠ 0, y′(t) = 0.

Turning points of a parallel curve are easily found from equation (9.32).

**9.7.2. Theorem.** The turning points, inflection points, and vertices of the parallel curve $p_d(t)$ correspond to those of the original curve p(t) (except when $\kappa(t) = -1/d$ at a turning point or vertex of p(t), in which case the corresponding point on $p_d(t)$ is then a cusp or an extraordinary point).

**Proof.** The theorem follows from equations (9.32)–(9.34).

Turning points depend on the coordinate system. On the other hand, inflection points and vertices are intrinsic properties of the curve. All three are invariant under regular reparameterizations.

We finish this section by listing some facts about length and area for parallel curves.

**9.7.3. Theorem.** If p(t), t ∈ [0,1], is a regular plane curve of total length L, then the total length $L_d$ of the nondegenerate parallel curve $p_d(t)$ is given by

$$L_d = |L + d\,\Delta\theta|,$$

where Δθ is the total angle of rotation of the principal normal n(t) of p(t). If the curve corresponding to p(t) is closed and convex, then

$$L_d = L + 2\pi|d|.$$

**Proof.** See [FarN90a] for the first part.

**9.7.4. Theorem.** If p(t), t ∈ [0,1], is a regular plane curve of total length L, then the area A between p(t) and the nondegenerate parallel curve $p_d(t)$ is given by

$$A = \frac{1}{2}(L + L_d)|d|,$$

where $L_d$ is the length of $p_d(t)$. If the curve corresponding to $p(t)$ is closed and convex, then

$$A = L|d| + \pi d^2.$$

**Proof.** See [FarN90a].

One can relate parallel curves and evolutes to the Gauss map. See [BaGM82].

## 9.8 Metric Properties of Surfaces

Now that we know a little about the metric properties of curves in $\mathbf{R}^n$, we move on to metric properties of Riemannian manifolds in general and surfaces in particular. We need to generalize what we did in $\mathbf{R}^n$, that is, our "local" view needs to be generalized to a global one. Probably the most natural first use of the metric is to define the length of curves in such manifolds and the notion of volume (area for surfaces). We start with the length of a curve in a manifold.

Let $\gamma:[a,b] \to \mathbf{M}^n$ be a curve in an n-dimensional submanifold $\mathbf{M}^n$ of $\mathbf{R}^k$. Recall that, as a curve in $\mathbf{R}^k$, the length of $\gamma$, $l_a^b\,\gamma$, is just the integral

$$\int_a^b |\gamma'|.$$

The only thing new will be how we interpret this formula. Note how the length depends on the length of the tangent vectors of the curve. However, as a curve **in** $\mathbf{M}^n$, we do not want to think of the tangent vectors as vectors in $\mathbf{R}^k$, but as vectors in the tangent spaces of $\mathbf{M}^n$ that are invariants associated to a manifold. Therefore, the length of a curve in a manifold should be thought of as a function of the dot product in each tangent space for the manifold, so that it is determined by the Riemannian metric for the manifold. Of course, as submanifolds of $\mathbf{R}^k$ we shall always be using the Riemannian metric induced by the standard dot product on $\mathbf{R}^k$. However, this particular choice is irrelevant and the reader should be aware of the fact that anything we write down involving dot products would be valid for any other Riemannian metric. Recall our comments at the beginning of the chapter that one should think of manifolds in an intrinsic way and forget the engulfing space $\mathbf{R}^k$.

Once we have a notion of length of a curve in a manifold, we can define the distance between its points.

**Definition.** The *distance between two points* $\mathbf{p}$ and $\mathbf{q}$ in a connected Riemannian manifold $\mathbf{M}^n$, denoted by $\mathrm{dist}(\mathbf{p},\mathbf{q})$, is defined by

$$\mathrm{dist}(\mathbf{p},\mathbf{q}) = \inf\{l_a^b\gamma \mid \gamma:[a,b] \to \mathbf{M}^n \text{ is a curve from } \mathbf{p} \text{ to } \mathbf{q}\}.$$

It is not hard to show that the function $\mathrm{dist}(\mathbf{p},\mathbf{q})$ is a metric on the manifold $\mathbf{M}^n$. Since this metric was derived from a Riemannian metric, it is the latter that is fundamental to the metric properties of a manifold. In differential geometry the Riemannian metric is usually brought into the picture using the following older terminology:

**Definition.** The quadratic form $Q_I$ defined on the tangent space $T_{\mathbf{p}}(\mathbf{M}^n)$ of the manifold $\mathbf{M}^n$ with Riemannian metric $\bullet$ by

$$Q_I(\mathbf{v}) = \mathbf{v} \bullet \mathbf{v}$$

is called the *first fundamental form* of the manifold at $\mathbf{p}$.

We conclude the following:

**9.8.1. Theorem.** The metric properties of a manifold are completely determined by its first fundamental form.

Next, we want to bring parameterizations into our discussion, because this is how manifolds are defined intrinsically, and we do not want to rely on them being contained in some $\mathbf{R}^k$. Locally, a manifold $\mathbf{M}^n$ looks like (can be identified with) $\mathbf{R}^n$. Formally, such an identification corresponds to a parameterization. The question is, if we do our work in $\mathbf{R}^n$, how can one use the parameterization to translate the results back to $\mathbf{M}^n$? We begin by answering this question for lengths of curves.

Let $\mathbf{U}$ be an open set in $\mathbf{R}^n$ and let $\Phi(u_1, u_2, \ldots, u_n)$,

$$\Phi: \mathbf{U} \to \mathbf{V} \subseteq \mathbf{M}^n,$$

be a parameterization for a neighborhood $\mathbf{V}$ of a point $\mathbf{p}$ on the manifold $\mathbf{M}^n$ in $\mathbf{R}^k$. Assume that $\mathbf{U}$ contains the origin and that $\Phi(\mathbf{0}) = \mathbf{p}$. It is easy to see that every curve $\gamma(t)$ in $\mathbf{V}$ through $\mathbf{p}$ can be expressed in the form

$$\gamma(t) = \Phi(\mu(t)),$$

where

$$\mu: (-a,a) \to \mathbf{R}^n \quad \text{and} \quad \mu(0) = \mathbf{0}.$$

See Figure 9.16. How does the length of $\mu$ relate to the length of $\gamma$?

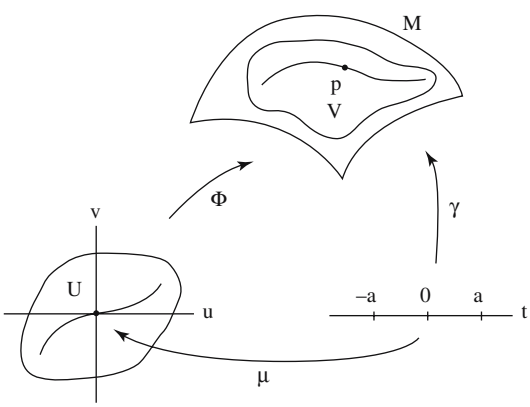



**Figure 9.16.** Representing curves in surfaces.

We know that the length of the curve $\gamma$ is computed from the lengths of its tangent vectors. If

$$\mu(t) = (\mu_1(t), \mu_2(t), \ldots, \mu_n(t)),$$

then the chain rule implies that

$$\gamma'(t) = \sum_{i=1}^{n} \frac{\partial \Phi}{\partial u_i}(\mu(t))\, \mu_i'(t). \tag{9.35}$$

**Definition.** The *metric coefficients* $g_{ij}$ of the parameterization $\Phi$ and their determinant g are the functions defined by

$$g_{ij} = \frac{\partial \Phi}{\partial u_i} \bullet \frac{\partial \Phi}{\partial u_j} \quad \text{and} \quad g = \det(g_{ij}).$$

Using the notation of metric coefficients, it follows from Equation (9.35) that

$$\gamma'(t) \bullet \gamma'(t) = \sum_{i,j=1}^{n} g_{ij}(\mu(t))\, \mu_i'(t)\mu_j'(t). \tag{9.36}$$

In other words, the length of the curve $\gamma(t)$ in the manifold is just the length of the corresponding ordinary curve $\mu(t)$ in $\mathbf{R}^n$ modified by the metric coefficients that depend on the parameterization and the Riemannian metric.

A vector $\mathbf{v}$ in the tangent plane at a point $\mathbf{p}$ of a manifold is just the tangent vector to some curve in the manifold. Therefore, if $\mathbf{v} = \gamma'(0)$, then we can use equation (9.36) to rephrase the definition of the fundamental form $Q_I(\mathbf{v})$ at $\mathbf{p}$ as

$$Q_I(\mathbf{v}) = Q_I(\gamma'(0)) = \sum_{i,j=1}^{n} g_{ij}(\mu(0))\, \mu_i'(0)\mu_j'(0). \tag{9.37}$$

**Definition.** The metric coefficients $g_{ij}$ are also called the *coefficients of the first fundamental form*.

Clearly, the matrix $(g_{ij})$ of metric coefficients with respect to a parameterization $\Phi$ is just the matrix of the symmetric bilinear map associated to the quadratic form $Q_I$ with respect to the basis consisting of the vectors $\partial\Phi/\partial u_i$ and g is the discriminant of $Q_I$ with respect to that basis. Because we have a positive definite form, it follows that $g > 0$ (Corollary 1.9.13).

In the case of a surface $\mathbf{S}$ the coefficients of the first fundamental form have historically been given the names E, F, and G, that is,

$$E(u,v) = g_{11}(u,v), \quad F(u,v) = g_{12}(u,v), \quad \text{and} \quad G(u,v) = g_{22}(u,v), \tag{9.38}$$

so that, letting $\mu(t) = (u(t), v(t))$, equation (9.37) is often written as

$$Q_I(\gamma') = E\,u'^2 + 2F\,u'v' + G\,v'^2. \tag{9.39}$$

**9.8.2. Example.** To compute E, F, G for the parameterization $\Phi(u,v) = \mathbf{p}_0 + u\mathbf{w}_1 + v\mathbf{w}_2$, $(u,v) \in \mathbf{R}^2$, of the plane through the point $\mathbf{p}_0$ with orthonormal basis $\mathbf{w}_1$ and $\mathbf{w}_2$.

**Solution.** Now

$$\frac{\partial\Phi}{\partial u} = \mathbf{w}_1 \quad \text{and} \quad \frac{\partial\Phi}{\partial v} = \mathbf{w}_2$$

implies that

$$g_{11} = 1, \quad g_{12} = g_{21} = 0, \quad \text{and} \quad g_{22} = 1.$$

Therefore, E = 1, F = 0, and G = 1.

**9.8.3. Example.** To compute E, F, G for the parameterization

$$\Phi(u,v) = (\cos u, \sin u, v), \quad (u,v) \in [0,2\pi] \times \mathbf{R},$$

of the cylinder $x^2 + y^2 = 1$.

**Solution.** This time

$$\frac{\partial\Phi}{\partial u} = (-\sin u, \cos u, 0) \quad \text{and} \quad \frac{\partial\Phi}{\partial v} = (0,0,1),$$

so that

$$g_{11} = 1, \quad g_{12} = g_{21} = 0, \quad \text{and} \quad g_{22} = 1.$$

Therefore, E = 1, F = 0, and G = 1. It may seem a little strange that the metric coefficients of the cyclinder are the same as those of a plane, but we shall see why this is as it should be shortly.

**9.8.4. Example.** To compute E, F, G for the spherical coordinate parameterization

$$\Phi(\theta,\phi) = (r\cos\theta\sin\phi, r\sin\theta\sin\phi, r\cos\phi), \quad (\theta,\phi) \in [0,2\pi] \times [0,\pi],$$

of the sphere of radius r about the origin.

**Solution.** We have

$$\frac{\partial\Phi}{\partial\theta} = (-r\sin\theta\sin\phi, r\cos\theta\sin\phi, 0) \quad \text{and} \quad \frac{\partial\Phi}{\partial\phi} = (r\cos\theta\cos\phi, r\sin\theta\cos\phi, -r\sin\phi).$$

Therefore, $E = r^2\sin^2\phi$, F = 0, $G = r^2$.

Equation (9.39) looks rather messy. If we write $\gamma(t) = (u(t),v(t))$, then Equation (9.39) is usually written in "differential" form as

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2.$$

Contrast this equation with the one that would be written down for the ordinary Euclidean plane:

$$ds^2 = du^2 + dv^2.$$

One sees that what is new is the metric coefficients associated to the surface **S**.

**Note.** We are **not** dealing with the **wedge** product of differential forms here as defined in Sections 4.9, 4.9.1, and 8.12. We are trying to specify a quadratic form, or equivalently, a bilinear map. A bilinear map <,> for the tangent spaces of an n-manifold corresponds to a tensor of type (0,2) and over a coordinate neighborhood $(\mathbf{U},\varphi)$ it can be expressed in the form

$$<,> = \sum_{i,j=1}^{n} g_{ij}\, du_i \otimes du_j.$$

See Section C.6. In other words, the expressions $du^2$, dudv, etc., above are abbreviations for $du\otimes du$, $du\otimes dv$, etc., respectively. It is unfortunate that similar notation is used but this is a historical legacy.

Summarizing our results for surfaces,

$$g = EG - F^2 > 0$$

and

$$\begin{aligned} l_0^t\gamma &= \int_0^t |\gamma'(t)|dt = \int_0^t \sqrt{Q_I(\gamma'(t))}dt \\ &= \int_0^t \sqrt{Eu'^2 + 2Fu'v' + Gv'^2}. \end{aligned}$$

Furthermore, the angle $\alpha$ between the two curvilinear coordinate axes

$$t \to \Phi(t,0) \quad \text{and} \quad t \to \Phi(0,t)$$

is given by

$$\cos\alpha = \frac{\dfrac{\partial\Phi}{\partial u} \bullet \dfrac{\partial\Phi}{\partial v}}{\left|\dfrac{\partial\Phi}{\partial u}\right|\left|\dfrac{\partial\Phi}{\partial v}\right|} = \frac{F}{\sqrt{EG}}.$$

Note that $EG > 0$ because $g > 0$.

Next, we show how to define and compute area for surfaces. Section 8.12 already gave an intrinsic definition of volume for an abstract oriented Riemannian manifold based on the theory of differential forms. We could build on that but differential forms are rather abstract. This chapter has mostly been following the more classical approach and we shall continue that to get some well-known formulas. We shall stick with manifolds defined by parameterizations and only require the reader to be familiar with basic advanced calculus. When appropriate we shall make some comments to tie what we do here to what was done in Section 8.12.

We indulge in one bit of generalization. Rather than restricting ourselves to parameterized surfaces in $\mathbf{R}^3$ we shall consider parameterized n-dimensional manifolds $\mathbf{M}^n$ in $\mathbf{R}^{n+1}$. Let $\Phi : \mathbf{U} \to \mathbf{M}$ be a regular parameterization. The parameterization $\Phi(u_1, u_2, \ldots, u_n)$ induces an orientation of $\mathbf{M}$. Assume that $\mathbf{n}(\mathbf{p})$ is the standard unit normal vector field on $\mathbf{M}$ determined by this orientation of $\mathbf{M}$. See Section 8.5.

**Definition.** The *volume* of the parameterization $\Phi$, denoted by $V(\Phi)$, is defined by

$$V(\Phi) = \int_{\mathbf{U}} \det \begin{pmatrix} \dfrac{\partial \Phi}{\partial u_1} \\ \vdots \\ \dfrac{\partial \Phi}{\partial u_n} \\ \mathbf{n} \end{pmatrix}.$$

When $n = 1$, volume is called *length*. When $n = 2$, volume is called *area*.

Note that $V(\Phi)$ is positive, since the normal vector was chosen in such a way as to make the determinant positive. Also, the volume might be infinite.

It is not hard to show that if $n = 1$, then this definition agrees with the definition of length of a curve given earlier. See Exercise 9.8.1 To justify this definition for arbitrary n, observe that the determinant is just the volume of the $(n + 1)$-dimensional parallelotope spanned by the $\partial\Phi/\partial u_i$ and $\mathbf{n}$ (Corollary 4.8.9). Since the normal vector $\mathbf{n}$ is orthogonal to the tangent plane and has length 1, the determinant is also the volume of the n-dimensional base of this parallelotope, namely, the parallelotope in the tangent plane spanned by the tangent vectors $\partial\Phi/\partial u_i$. See Figure 9.17. Therefore, to take the case of a surface as an example, the integral is just the limit of Riemann sums of areas of parallelograms that approximate the manifold.

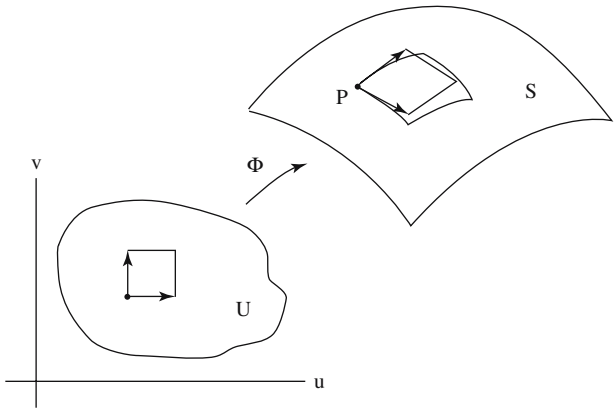
The definition of volume is not very satisfactory as it stands because the formula also involves the normal vector $\mathbf{n}$.

**9.8.5. Theorem.** Let $\Phi : \mathbf{U} \to \mathbf{M}^n$ be a regular parameterization of manifold $\mathbf{M}^n$ in $\mathbf{R}^{n+1}$. Then

$$V(\Phi) = \int_{\mathbf{U}} \sqrt{\det(g_{ij})} = \int_{\mathbf{U}} \sqrt{g}.$$

**Proof.** We have

**Figure 9.17.** Computing the area of a surface.




$$\det{}^2\begin{pmatrix} \dfrac{\partial\Phi}{\partial u_1} \\ \vdots \\ \dfrac{\partial\Phi}{\partial u_n} \\ \mathbf{n} \end{pmatrix} = \det\left(\begin{pmatrix} \dfrac{\partial\Phi}{\partial u_1} \\ \vdots \\ \dfrac{\partial\Phi}{\partial u_n} \\ \mathbf{n} \end{pmatrix}\begin{pmatrix} \dfrac{\partial\Phi^T}{\partial u_1} & \cdots & \dfrac{\partial\Phi^T}{\partial u_n}\mathbf{n}^T \end{pmatrix}\right)$$

$$= \det\begin{pmatrix} & & & 0 \\ & \dfrac{\partial\Phi}{\partial u_i}\bullet\dfrac{\partial\Phi}{\partial u_j} & & \vdots \\ & & & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

$$= \det(g_{ij}).$$

The second inequality followed from the fact that $\mathbf{n}$ is orthogonal to all the $\partial\Phi/\partial u_i$ and has unit length.

Theorem 9.8.5 shows explicitly how the first fundamental form determines another metric quantity, namely, volume. Keep in mind though that we do not yet have a notion of the volume of a manifold, but only a notion of the volume of a parameterization. The next theorem shows that we are actually dealing with an invariant of the manifold.

**9.8.6. Lemma.** If $\Phi_1$ and $\Phi_2$ are two regular one-to-one parameterizations of an n-dimensional manifold $\mathbf{M}^n$ in $\mathbf{R}^{n+1}$ that induce the same orientation on $\mathbf{M}$, then

$$V(\Phi_1) = V(\Phi_2).$$

**Proof.** This is an easy consequence of the change of variable theorem for multiple integrals and the fact that $\Phi_2$ is a reparameterization of $\Phi_1$. We need to assume that the parameterizations induce the same orientations in order to be able to guarantee that the reparameterization map has a positive determinant. Otherwise, the sign of the volumes may change.

**Definition.** Let $\mathbf{M}^n$ be an oriented compact n-dimensional submanifold of $\mathbf{R}^{n+1}$ that admits a regular one-to-one parameterization $\Phi$ that induces the same orientation. Define the *volume* of $\mathbf{M}$, denoted by volume ($\mathbf{M}$), by

$$\text{volume}(\mathbf{M}) = V(\Phi).$$

When n = 1 or 2, one uses the terms *length* and *area*, respectively, instead of the generic term "volume."

We need compactness in the definition to guarantee that the volume will be finite. An immediate consequence of Lemma 9.8.6 is that the definition of volume of a manifold does not depend on the parameterization.

**9.8.7. Theorem.** The volume of an oriented compact n-dimensional submanifold $\mathbf{M}^n$ of $\mathbf{R}^{n+1}$ that admits a regular one-to-one parameterization is well defined.

Compare Theorem 9.8.5, the formula for our current definition of volume, with Theorem 8.12.19, which is the differential form version. Roughly speaking, Theorem 9.8.5 is Theorem 8.12.19 for one coordinate neighborhood $(\mathbf{M},\Phi^{-1})$. Volume, as defined in Section 8.12, was by definition an intrinsic property of a manifold. Our current approach needed Lemma 9.8.6 to get the same result. Exercise 9.8.2 asks the reader to show that this new definition again agrees with the definition given in Section 9.2 when n = 1, that is, in the case of curves in the plane.

Specializing to surfaces in $\mathbf{R}^3$, we get

**9.8.8. Theorem.** Let $\Phi:\mathbf{U} \to \mathbf{S}$ be a regular and one-to-one parameterization of a surface $\mathbf{S}$ in $\mathbf{R}^3$. Then

$$\text{area}(\mathbf{S}) = \iint_{\mathbf{U}} \sqrt{EG - F^2}.$$

**Proof.** This is an immediate consequence of Theorem 9.8.5 and the relation between integrals over regions in the plane and double integrals.

Note the important identity

$$|\Phi_u \times \Phi_v|^2 = (\Phi_u \bullet \Phi_u)(\Phi_v \bullet \Phi_v) - (\Phi_u \bullet \Phi_v)^2 = EG - F^2, \tag{9.40}$$

which follows from Proposition 1.10.4(3).

**9.8.9. Example.** To find the area of a region $\mathbf{S}$ in the plane $\mathbf{R}^2$.

**Solution.** If we use the identity map to parameterize $\mathbf{S}$, then it is easy to check that g = 1 and

$$\text{area}(\mathbf{S}) = \iint_{\mathbf{S}} 1,$$

which is the usual calculus definition of area.

The definitions above are inadequate for finding the area or volume of even simple manifolds because the natural parameterizations are often neither regular nor one-

to-one. For example, natural parameterizations of a cylinder or sphere (see the examples below) are only regular and/or one-to-one on the interior of their domains. However, the definition of volume and the formulas for computing it easily extend in such special cases because they have the property that the boundary of the domain is a set of measure 0 and the parameterization is regular and one-to-one on the interior of the domain. The proof would involve a straightforward limit process. We shall use such extended formulas in the examples below. The reader who is uncomfortable with this should compute areas for slightly smaller regions in the surface. Finally, although we will not do this here, it should be pointed out that one **can** also define the volume of an **arbitrary** n-dimensional submanifold of $\mathbf{R}^{n+1}$. If there is no single parameterizing function, then the definition is based on piecing together the volumes defined for nice local parameterizations like in the case for length.

**9.8.10. Example.** To find the area of the bounded cylinder

$$\mathbf{C} = \{(x,y,z) \mid x^2 + y^2 = 1 \text{ and } 0 \le z \le 1\},$$

**Solution.** Using the parameterization

$$\Phi(u,v) = (\cos u, \sin u, v), \quad (u,v) \in [0,2\pi] \times [0,1],$$

and the computations made in Example 9.8.3, we see that g = 1 and by Theorem 9.8.8

$$\text{area}(\mathbf{C}) = \iint_{[0,2\pi]\times[0,1]} 1 = 2\pi.$$

**9.8.11. Example.** To find the area of the sphere **S** of radius r about the origin.

**Solution.** We again use a spherical coordinate parameterization

$$\mathbf{\Phi}(\theta,\phi) = (r\cos\phi\sin\phi, r\sin\theta\sin\phi, r\cos\phi), \quad (\theta,\phi) \in [0,2\pi] \times [0,\pi].$$

Now, Example 9.8.4 showed that $E = r^2\sin^2\phi$, $F = 0$, $G = r^2$. Therefore,

$$EG - F^2 = r^4 \sin^2 \phi$$

and

$$\text{area}(\mathbf{S}) = \iint_{[0,2\pi]\times[0,\pi]} r^2 \sin\phi = 4\pi r^2.$$

We finish this section with some comments on the extent to which the results above are influenced by the parameterization that was chosen.

First of all, one could define the first fundamental form (or Riemannian metric) by means of a parameterization and equation (9.37). It will be a good exercise in the use of the chain rule and the definition of tangent vectors to show directly that such a definition would be independent of the parameterizations. Let $\mathbf{U}, \mathbf{V} \subseteq \mathbf{R}^n$ and let

$$\Psi: \mathbf{V} \to \mathbf{M}^n \quad \text{and} \quad \Phi: \mathbf{U} \to \mathbf{M}^n$$

be two parameterizations of an n-dimensional manifold $\mathbf{M}^n$ in $\mathbf{R}^k$. Assume that $\Phi = \Psi\circ\sigma$, where $\sigma : \mathbf{U} \to \mathbf{V}$, $\sigma(x_1,x_2, \ldots , x_n) = (u_1,u_2, \ldots , u_n)$, is a change of coordinate transformation. Let $Q_I(\Psi)$ and $Q_I(\Phi)$ denote the first fundamental forms of $\Psi$ and $\Phi$, respectively. Using the notation in equations (9.35–37), let $\lambda = \sigma\circ\mu$. Then the previous equations and a generous use of the chain rule shows that

$$\begin{aligned}
Q_1(\Phi)(\gamma') &= \sum_{i,j=1}^{n} \frac{\partial\Phi}{\partial x_i} \bullet \frac{\partial\Phi}{\partial x_j} \mu_i'\mu_j' \\
&= \sum_{i,j=1}^{n} \left[ \left( \sum_{s=1}^{n} \frac{\partial\Psi}{\partial\sigma_s} \frac{\partial\sigma_s}{\partial x_i} \right) \bullet \left( \sum_{t=1}^{n} \frac{\partial\Psi}{\partial\sigma_t} \frac{\partial\sigma_t}{\partial x_j} \right) \right] \mu_i'\, \mu_j' \\
&= \sum_{s,t=1}^{n} \frac{\partial\Psi}{\partial x_s} \bullet \frac{\partial\Psi}{\partial x_t} \lambda_s'\lambda_t' \\
&= Q_1(\Psi)(\lambda').
\end{aligned}$$

This is what we wanted to show.

Volume also was independent of parameterizations. The metric coefficients on the other hand are not but the next theorem shows that they transform in a well-defined way.

**9.8.12. Theorem.** In the case of surfaces the metric coefficients transform as follows (using the notation above and writing $(u,v) = \sigma(x,y)$):

$$\begin{aligned}
E(\Phi) &= E(\psi)u_x^2 + 2F(\psi)u_x v_x + G(\psi)v_x^2 \\
F(\Phi) &= E(\psi)u_x u_y + F(\psi)(u_x v_y + v_x u_y) + G(\psi)v_x v_y \\
G(\Phi) &= E(\psi)u_y^2 + 2F(\psi)u_y v_y + G(\psi)v_y^2
\end{aligned}$$

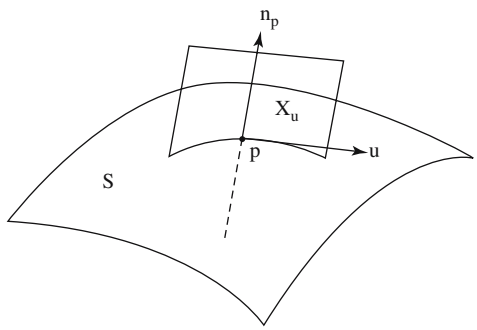
**Proof.** One again simply uses the appropriate chain rule. See [Lips69].

## 9.9 The Geometry of Surfaces

Our overview of the geometry of surfaces in $\mathbf{R}^3$ will start with a sketch of the historical development of a few of the really fundamental invariants that are central to any understanding of surfaces before we develop the results rigorously using modern terminology. It should not be surprising that the initial attempt to study the geometry of surfaces was by means of curves. The first such result was due to Euler and dates from 1760.

Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and $\mathbf{p}$ any point of $\mathbf{S}$. Let $\mathbf{n_p}$ be a unit vector normal to $\mathbf{S}$ at $\mathbf{p}$. For each unit vector $\mathbf{u}$ in the tangent plane to $\mathbf{S}$ at $\mathbf{p}$, let $\mathbf{X_u}$ be the plane through $\mathbf{p}$ generated by the vectors $\mathbf{u}$ and $\mathbf{n_p}$. Consider the curve that is the intersection of $\mathbf{X_u}$ with $\mathbf{S}$. Such a curve is called a *normal section* of $\mathbf{S}$ at $\mathbf{p}$. Let $\boldsymbol{\kappa_u}$ denote the signed curvature of this curve, where the plane $\mathbf{X_u}$ is given the orientation induced by the ordered basis $(\mathbf{u},\mathbf{n_p})$. See Figure 9.18.

**Figure 9.18.** Normal curvatures at a point.




**Euler's Theorem.** As $\mathbf{u}$ ranges over all unit vectors in the tangent plane to $\mathbf{S}$ at $\mathbf{p}$, the set of curvatures $\kappa_{\mathbf{u}}$, called *normal curvatures*, forms a closed interval $[\kappa_1, \kappa_2]$. If the $\kappa_{\mathbf{u}}$ are not all equal, then there is precisely one direction $\mathbf{u}_1$ for which the associated curve has the minimum curvature $\kappa_1$ and one direction $\mathbf{u}_2$ for which the associated curve has the maximum curvature $\kappa_2$. Furthermore, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal and if a unit vector $\mathbf{u}$ makes an angle $\theta$ with $\mathbf{u}_1$, then

$$\kappa_{\mathbf{u}} = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta. \tag{9.41}$$

**Definition.** The curvatures $\kappa_1$ and $\kappa_2$ are called the *principal normal curvatures* of $\mathbf{S}$ at $\mathbf{p}$. and the radii of the associated osculating circles are the *principal radii*. The vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are called the *principal normal directions* of $\mathbf{S}$ at $\mathbf{p}$.

Note that if we change the direction of the normal $\mathbf{n_p}$, then the sign of the curvatures $\kappa_{\mathbf{u}}$ changes, but all the results remain valid. Another interesting fact is that we really did not have to consider planar curves (curves that lie in $\mathbf{X_u}$), but any parametric curve in $\mathbf{S}$ through $\mathbf{p}$ that has tangent vector equal to $\mathbf{u}$ would give rise to the same curvature $\kappa_{\mathbf{u}}$.

Let $\mathbf{c_u}$ denote the center of curvature at the point $\mathbf{p}$ of the planar curve that was used to define the normal curvature $\kappa_{\mathbf{u}}$. The centers $\mathbf{c_u}$ lie on the line normal to $\mathbf{S}$ at $\mathbf{p}$, so that $\mathbf{c_u} = \mathbf{p} + t_{\mathbf{u}}\mathbf{n_p}$ for some $t_{\mathbf{u}} \in \mathbf{R}$. The set of these centers as $\mathbf{u}$ ranges over all unit vectors in the tangent plane consists of one or two closed intervals. More precisely, there exist numbers

$$t_{\min} \text{ and } t_{\max}, \quad -\infty \le t_{\min} \le t_{\max} \le \infty,$$

so that the set of centers is either

$$[\mathbf{c}_{\min}, \mathbf{c}_{\max}] \text{ or } [-\infty, \mathbf{c}_{\min}] \cup [\mathbf{c}_{\max}, \infty],$$

where $\mathbf{c}_{\min} = \mathbf{p} + t_{\min}\mathbf{n_p}$ and $\mathbf{c}_{\max} = \mathbf{p} + t_{\max}\mathbf{n_p}$. The points $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ are the centers of circles that define the principal normal curvatures $\kappa_1$ and $\kappa_2$, respectively. See Figure 9.19. The point $\mathbf{p}$ in Figure 9.19(a) is called an *elliptic point* of the surface because the surface lies to one side of the tangent plane and meets the tangent plane in a single point locally. The point $\mathbf{p}$ in Figure 9.19(b) is called a *hyperbolic point* because the

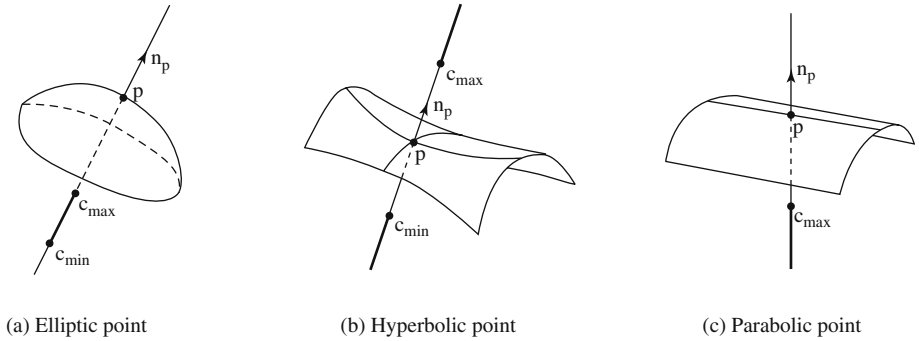



**Figure 9.19.** Local curvature properties of surfaces.

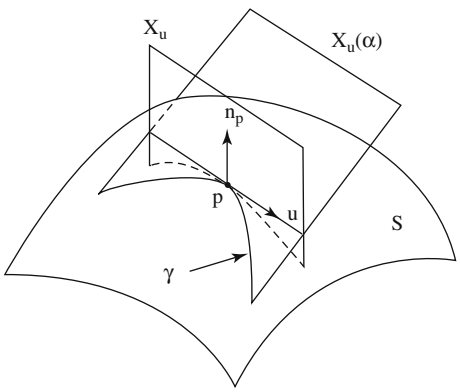



**Figure 9.20.** Meusnier's theorem.

surface lies on both sides of the tangent plane. The point **p** in Figure 9.19(c) is called a *parabolic point* because the surface lies to one side of the tangent plane and meets the tangent plane in a line locally.

**Definition.** The points $\mathbf{c}_{\min}$ and $\mathbf{c}_{\max}$ are called the *foci* of the normal line at **p** or the *focal points* of **S** at **p**. The set of focal points of **S** is called the *surface of centers*, the *focal surface*, or the *evolute* of **S**.

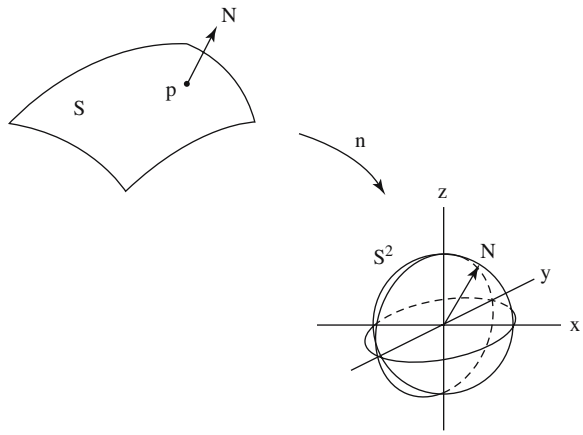
See [HilC99] or [Gray98] for more about focal points. We shall have more to say about the surface of centers and evolutes in Sections 9.12 and 9.13.

Euler's theorem was soon generalized in 1776 by Meusnier. Keeping the same notation as above, let $\mathbf{X_u}(\alpha)$, $0 \le \alpha < \pi/2$, be a plane through **p** that contains **u** and makes an angle $\alpha$ with $\mathbf{X_u}$. Consider the parametric curve $\gamma$ parameterized by arc-length that parameterizes the intersection of $\mathbf{X_u}(\alpha)$ and **S** in a neighborhood of **p** and which has tangent vector **u** at **p**. Let $\boldsymbol{\kappa}_{\mathbf{u}}(\alpha)$ denote the curvature of that curve at **p**. See Figure 9.20.

**Meusnier's Theorem.** $\boldsymbol{\kappa}_{\mathbf{u}}(\alpha)\cos\alpha = \boldsymbol{\kappa}_{\mathbf{u}}$.

Both Euler's and Meusnier's theorem will be proved shortly. Although interesting, these two theorems still deal with curves and are only indirectly about surfaces. It was

**Figure 9.21.** The Gauss map.




Gauss who, in 1827, presented the first truly intrinsic surface invariants. Let **S** be an **oriented** surface in $\mathbf{R}^3$. Recall that saying **S** is oriented is equivalent to saying that we have made a consistent choice of unit normals at each point of **S** so that the map

$$\mathbf{n}: \mathbf{S} \to \mathbf{S}^2,$$

where $\mathbf{n}(\mathbf{p})$ is the chosen unit normal at **p** in **S**, is a nice smooth map.

**Definition.** The map **n** is called the *Gauss map* for the **oriented** surface **S**.

See Figure 9.21. Compare this definition with what we called the Gauss map for plane curves in Section 9.3. The assumption that the surface is oriented is never a problem. First of all, we usually only need it for a small neighborhood of a point. Second of all, we will be dealing with parameterized surfaces in which case we can, and shall, always assume that the orientation is induced by the given parameterization.

Following Gauss we can now define a fundamentally important intrinsic invariant of surfaces.

**A geometric definition of Gauss curvature**. Let **n** be the Gauss map for some surface **S**. Then the value $\mathrm{K}(\mathbf{p})$ defined by

$$\mathrm{K}(\mathbf{p}) = \lim_{\mathbf{U}\to\mathbf{p}} \frac{\text{“signed area”}(\mathbf{n}(\mathbf{U}))}{\text{area}(\mathbf{U})}, \tag{9.42}$$

where **U** is a neighborhood of **p** in **S**, is called the *Gauss curvature* of **S** at **p**. By the term “signed area” in the numerator we mean that we set it to the positive or negative value of $\text{area}(\mathbf{n}(\mathbf{U}))$ depending on whether the map **n** is orientation preserving or not.

This definition is not very rigorous, although it could be made so. We shall give a better definition shortly, but it is a nice intuitive way to think about the concept.

**9.9.1. Example.** The Gauss curvature of a sphere of radius r is $1/r^2$. To justify this note that computing the area of a region on a sphere in spherical coordinates involves a similar integral to the one used in Example 9.8.11. The only difference is the domain of the integral, which would now be an arbitrary region in the $\theta$-$\phi$ plane. This domain would not change as we pass from of a region on a sphere of radius r to the projected region on the sphere of radius 1. The only difference is the factor $r^2$.

**9.9.2. Example.** The Gauss curvature of a plane is 0. The reason is that the normal of a plane is constant so that the Gauss map maps any region to a point that has zero area.

Here are some other intuitive observations about the Gauss curvature. If we look at an elliptic point (see Figure 9.19(a)) we see that the Gauss curvature is positive. On the other hand, at a hyperbolic point (see Figure 9.19(b)) the curvature is negative because the Gauss map traces out an opposite oriented area on the sphere. We shall see shortly that the Gauss curvature is the product of the principal curvatures (Theorem 9.9.9) and so it vanishes at parabolic points.

**9.9.3. Example.** The Gauss curvature of a cylinder is 0. The justification for this using only the definition of the Gauss map is that the normal is constant along the generating lines for a cylinder that implies that all regions on the cylinder are mapped into one-dimensional sets on the unit sphere and these clearly have zero area.

Does Example 9.9.3 mean that we have a wrong notion of curvature for a surface? Not really, but we will have to explain the sense in which we are using the word "curvature." Gauss curvature is a fundamental invariant of a surface, but before we analyze it further we digress briefly to describe an analogous notion for polygonal surfaces. This should help reinforce the geometric idea we are discussing.

Let **S** be a polygonal surface. It is convenient to assume in the discussion below that all of its facets are **triangles**. Clearly, any "curvature" of such a surface is concentrated at vertices. One might be tempted to think of there being curvature along an edge, but just as cylinders have zero Gauss curvature, so will points along edges end up having zero curvature. We need a few definitions before we are ready to define a concept of Gauss curvature in the polygonal case.

**Definition.** Let **S** be a surface in $\mathbf{R}^n$. The surface **S** is said to be *convex* if it is contained in the boundary of a convex polyhedron in $\mathbf{R}^n$. If $\mathbf{p} \in \mathbf{S}$, then **S** is *convex at the point* **p** if a neighborhood of the point **p** in **S** is contained in the boundary of a convex polyhedron in $\mathbf{R}^n$.

**Definition.** A set $\mathbf{X} \subseteq \mathbf{S}^2$ is said to be *geodesically convex* if for every two distinct points **p** and **q** in **X**, one of the two arcs of the great circle in $\mathbf{S}^2$ through **p** and **q** that connect **p** and **q** lies in **X**. The *geodesic convex closure* of **X** is defined to be the intersection of all geodesically convex sets in $\mathbf{S}^2$ that contain **X** (intuitively, it is the "smallest" geodesically convex set containing **X**).

Geodesic convexity and geodesic convex closure are the smooth-surface analogs of convex and convex closure for vector spaces. The reason for the term "geodesic" in

the definition will become clear after we define geodesic curves in the next section, since arcs of great circles are geodesic curves. Clearly, the sphere and any hemisphere in it are geodesically convex.

**Definition.** Let $\mathbf{p}$ be a point of a polygonal surface $\mathbf{S}$ in $\mathbf{R}^3$. Assume that $\mathbf{S}$ is oriented, so that we have a notion of "outward-pointing" normal vectors for its facets. Let $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_k \in \mathbf{S}^2$ be the outward pointing unit normal vectors to the facets of $\mathbf{S}$ that contain $\mathbf{p}$. The *normal neighborhood* of $\mathbf{p}$ in $\mathbf{S}$ is defined to be the subset of $\mathbf{S}^2$ that is the geodesic convex closure of the set of points $\{\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_k\}$ in $\mathbf{S}^2$. The *normal angle* of $\mathbf{p}$ is the area of the normal neighborhood of $\mathbf{p}$.

**Note.** In classical solid geometry the area of a subset $\mathbf{X}$ of $\mathbf{S}^2$ is called the *measure* of the *solid angle subtended by* $\boldsymbol{X}$.

For example, the normal neighborhood of a vertex of a cube is a curved triangle which fills an eighth of a sphere and has normal angle $\pi/2$. More generally, in Figure 9.22, $\mathbf{n_A}$, $\mathbf{n_B}$, $\mathbf{n_C}$, and $\mathbf{n_D}$ are the normals to the triangles $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ in Figure 9.22(a), respectively. The shaded region in Figure 9.22(c) is the normal neighborhood of the vertex $\mathbf{p}$ in Figure 9.22(a) and is the geodesic convex hull of the normals $\mathbf{n_A}$, $\mathbf{n_B}$, $\mathbf{n_C}$, and $\mathbf{n_D}$. In general, if $\mathbf{n}_i$ is the normal to the facet $\mathbf{f}_i$, then the arc from $\mathbf{n}_i$ to $\mathbf{n}_{i+1}$ consists of the normals of the planes one gets by rotating the plane containing $\mathbf{f}_i$ to the one containing $\mathbf{f}_{i+1}$ about the common edge keeping both facets on the same side of the planes. It is easy to see that such normals generate an arc of a great circle. If the surface is convex at the vertex, then one can also show that the normal neighborhood consists of all the normals of the support planes to the surface at that vertex. (A *support plane* at a point $\mathbf{p}$ of a set $\mathbf{X}$ in $\mathbf{R}^n$ is a hyperplane containing $\mathbf{p}$ with the property that $\mathbf{X}$ lies in one of the halfplanes defined by the hyperplane.)

We can now define the Gauss curvature at a point of $\mathbf{S}$.

**Definition.** Let $\mathbf{p}$ be a point of a polygonal surface $\mathbf{S}$ in $\mathbf{R}^3$. Define the *polygonal Gauss curvature* function $K_p$ on $\mathbf{S}$ by

$$K_p(\mathbf{p}) = \frac{\text{normal angle at } \mathbf{p} \text{ in } \mathbf{S}}{\text{area of neighborhood } \mathbf{U} \text{ of } \mathbf{p} \text{ in } \mathbf{S}},$$

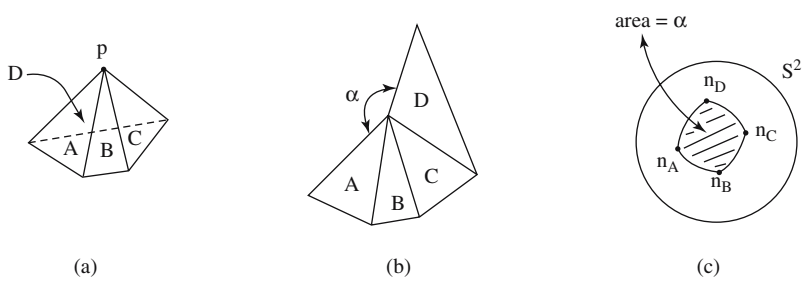



**Figure 9.22.** Gauss curvature for polygonal manifolds.

where **U** is chosen in some appropriate consistent manner as we move from point to point. (For example, Calladine ([Call86]) choses **U** to have one third of the area of all the facets adjacent to **q**.)

The validity of an earlier comment is now clear. The polygonal Gauss curvature vanishes at nonvertex points. If the point lies in the interior of a facet, then the normal neighborhood is just a point. If the point lies in the interior of an edge, then the normal neighborhood is just an arc of a great circle. In both cases, the area of the neighborhoods is zero.

Curvature is related to another important quantity in the polygonal context.

**Definition.** The *angular defect* at a point **p** in **S** is defined to be 0 if **p** is not a vertex and if **p** is a vertex, then it is

$$2\pi - (\text{sum of the interior angles at } \mathbf{p} \text{ of the facets of } \mathbf{S} \text{ that meet at the } \mathbf{p}).$$

See Figure 9.22 again. The angular defect at the vertex **p** in Figure 9.22(a) is the angle $\alpha$ shown in Figure 9.22(b). The angular defect clearly has some relation to curvature, because the larger it is, the more pointed the surface is at a vertex.

**9.9.4. Theorem.** Let **p** be a point of a polygonal surface **S** in $\mathbf{R}^3$. Then

$$\text{angular defect at } \mathbf{p} \text{ in } \mathbf{S} = \text{normal angle of } \mathbf{p} \text{ in } \mathbf{S}.$$

**Proof.** The only case that has to be considered is where **p** is a vertex. See Figure 9.22. The angle $\alpha$ in Figure 9.22(b) is the angular defect at the vertex **p** in Figure 9.22(a). The area of the shaded region in Figure 9.22(c) is the normal angle of **p**. See [Call86], [Crom97], or [HilC99] for a proof. A basic element is the classical Greek theorem that states that the area of a region in a sphere that is bounded by arcs of great circles depends only on sum of the exterior angles of the region.

A consequence of Theorem 9.9.4 is worth noting, namely, the normal angle at a vertex is unchanged if the surface is "folded" arbitrarily at the vertex by deformations that move the facets rigidly changing the "creases" (the angles between the faces along a common edge coming out of the vertex). This is because the angular defect does not change. When facets are triangles and there are more than three at the vertex, then many different folds are possible. [Call86] describes some applications of this result.

The discussion up to this point in this section was intended to motivate some important concepts about surfaces and was not entirely rigorous. We shall now consider smooth surfaces again and redo some of what we have covered but will use definitions that permit a more rigorous development. First of all, returning to the Gauss map $\mathbf{n}(\mathbf{p})$, note that its derivative $D\mathbf{n}(\mathbf{p})$ at a point **p** of **S** maps the tangent space $T_{\mathbf{p}}(\mathbf{S})$ to the tangent space to $\mathbf{S}^2$ at $\mathbf{n}(\mathbf{p})$. Since the two tangent spaces are the same (the tangent planes are parallel), we consider $D\mathbf{n}(\mathbf{p})$ to be a linear map

$$D\mathbf{n}(\mathbf{p}) : T_{\mathbf{p}}(\mathbf{S}) \to T_{\mathbf{p}}(\mathbf{S}).$$

**Definition.** The map $D\mathbf{n}(\mathbf{p})$ is called the *Weingarten map*.

**Note.** We have given different definitions of surfaces and manifolds and their tangent spaces. Therefore, to avoid any confusion on the part of the reader when it

comes to computations, we again restate our assumption that in this chapter we are following the definitions in Section 8.4. Tangent spaces are considered to be defined in terms of tangent vectors to curves and derivatives map tangent vectors to tangent vectors of the composite curves. For example, in case of the Gauss map, if $\mathbf{v}$ is a vector in $T_{\mathbf{p}}(\mathbf{S})$ and $\gamma(t)$ is a curve in $\mathbf{S}$ with $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$, then

$$D\mathbf{n}(\mathbf{p})(\mathbf{v}) = (\mathbf{n} \circ \gamma)'(0). \tag{9.43}$$

**9.9.5. Lemma.** The Weingarten map $D\mathbf{n}(\mathbf{p})$ is a self-adjoint linear map.

**Proof.** One needs to show that

$$D\mathbf{n}(\mathbf{p})(\mathbf{v}_1) \bullet \mathbf{v}_2 = \mathbf{v}_1 \bullet D\mathbf{n}(\mathbf{p})(\mathbf{v}_2)$$

for an arbitrary basis $(\mathbf{v}_1, \mathbf{v}_2)$ of $T_{\mathbf{p}}(\mathbf{S})$. This involves fairly straightforward computations using the definitions. See [DoCa76].

Lemma 9.9.5 is a technical fact, which, along with the next two definitions, will lead to some geometric results.

**Definition.** The quadratic form $Q_{II}$ on $T_{\mathbf{p}}(\mathbf{S})$ defined by

$$Q_{II}(\mathbf{v}) = -\, D\mathbf{n}(\mathbf{p})(\mathbf{v}) \bullet \mathbf{v}$$

is called the *second fundamental form* of $\mathbf{S}$ at $\mathbf{p}$.

The minus sign is used to reduce number of minus signs elsewhere. The second fundamental form is really a directional derivative. It measures the turning of the tangent plane in the direction defined by a tangent vector.

**Definition.** The map

$$S_{\mathbf{p}} = -D\mathbf{n}(\mathbf{p}) : \; T_{\mathbf{p}}(\mathbf{S}) \to T_{\mathbf{p}}(\mathbf{S})$$

is called the *shape operator* for the surface $\mathbf{S}$.

Some authors (see [ONei66]) make the latter the basis of the study of surfaces. It clearly contains the same information as the second fundamental form. If the point $\mathbf{p}$ is arbitrary, we shall drop the subscript $\mathbf{p}$ in the notation for the shape operator and simply write S (in the same sense that one could simply write $D\mathbf{n}$). For example, the notation $S(\mathbf{v}) \bullet \mathbf{v}$ is an abbreviation for the statements $S_{\mathbf{p}}(\mathbf{v}) \bullet \mathbf{v}$ for $\mathbf{p} \in \mathbf{S}$ with $\mathbf{v}$ the appropriate vector in $T_{\mathbf{p}}(\mathbf{S})$.

**Definition.** Let $\gamma(t)$ be a regular curve in $\mathbf{S}$ passing through $\mathbf{p}$. Let $\mathbf{N}$ be the principal normal and the $\boldsymbol{\kappa}$ curvature of $\gamma$ at $\mathbf{p}$. Let $\theta$ be the angle between $\mathbf{N}$ and $\mathbf{n}$ at $\mathbf{p}$. The number

$$\boldsymbol{\kappa}_{n,\gamma}(\mathbf{p}) = \boldsymbol{\kappa} \cos\theta = \boldsymbol{\kappa}\, \mathbf{N} \bullet \mathbf{n}(\mathbf{p})$$

is called the *normal curvature* of $\gamma$ at $\mathbf{p}$ in $\mathbf{S}$.

We analyze this definition of normal curvature further. Let $\gamma(s)$ be a (regular) curve lying in the surface $\mathbf{S}$ parameterized by arc-length and let $\mathbf{N}(s)$ be its principal normal.

Assume that $\gamma(0) = \mathbf{p} \in \mathbf{S}$. First of all, since $\gamma'' = \boldsymbol{\kappa}\mathbf{N}$, the normal curvature $\boldsymbol{\kappa}_{n,\gamma}(\mathbf{p})$ is now also defined by the formula

$$\boldsymbol{\kappa}_{n,\gamma}(\mathbf{p}) = \gamma''(t) \bullet \mathbf{n}(\mathbf{p}).$$

Next, to simplify the notation, let $\mathbf{n}(s)$ be shorthand for $\mathbf{n}(\gamma(s))$. Then $\mathbf{n}(s) \bullet \gamma'(s) = 0$ and differentiating both sides of this identity gives

$$\mathbf{n}'(s) \bullet \gamma'(s) = -\mathbf{n}(s) \bullet \gamma''(s). \tag{9.44}$$

It follows that

$$\begin{aligned} Q_{II}(\gamma'(0)) &= -D\mathbf{n}(\mathbf{p})(\gamma'(0)) \bullet \gamma'(0) \\ &= -\mathbf{n}'(0) \bullet \gamma'(0) \\ &= \mathbf{n}(0) \bullet \gamma''(0) \qquad \text{(by Equation (9.44))} \\ &= \mathbf{n}(0) \bullet \boldsymbol{\kappa}(0)\mathbf{N}(0) \\ &= \boldsymbol{\kappa}_{n,\gamma}(\mathbf{p}), \end{aligned}$$

and we have shown

**9.9.6. Theorem.** The value of the second fundamental form at a vector $\mathbf{v}$ in the tangent space to a surface at some point $\mathbf{p}$ is the normal curvature of any regular curve in the surface through $\mathbf{p}$ with tangent vector $\mathbf{v}$ there.

Theorem 9.9.6 proves Meusnier's theorem, which we restate now in the following way:

**9.9.7. Theorem.** (Meusnier) All regular curves lying in a surface $\mathbf{S}$ passing through a point $\mathbf{p}$ and having the same tangent vector at $\mathbf{p}$ have the same normal curvatures there.

Note that by definition

$$Q_{II}(\gamma'(0)) = S_{\mathbf{p}}(\gamma'(0)) \bullet \gamma'(0).$$

Therefore, the normal curvature of a curve can be derived from the shape operator. In fact, by normalizing vectors, we get information about the curvature of the surface.

**Definition.** Let $\mathbf{u}$ be a unit vector in the tangent space to $\mathbf{S}$ at $\mathbf{p}$. Then the quantity

$$\boldsymbol{\kappa}_n(\mathbf{u}) = S_{\mathbf{p}}(\mathbf{u}) \bullet \mathbf{u}$$

is called the *normal curvature* of $\mathbf{S}$ *in the direction* $\mathbf{u}$.

**Note:** There will never be any confusion when using the term "normal curvature" at a point $\mathbf{p}$ of a surface $\mathbf{S}$. Although we have two variants, one, the normal curvature $\boldsymbol{\kappa}_{n,\gamma}$ of a curve $\gamma$, and the other, the normal curvature $\boldsymbol{\kappa}_n(\mathbf{u})$ in a direction defined by a

unit tangent vector $\mathbf{u}$, they have the same value when $\mathbf{u}$ is the tangent vector to $\gamma$ at $\mathbf{p}$. The only difference is that, as we have shown, $\kappa_{n,\gamma}$ is completely defined by the tangent vector to $\gamma$ at $\mathbf{p}$ and thus basically a function defined on **all** tangent vectors, whereas $\kappa_n$ is only defined on unit tangent vectors. For that reason one can use the symbols $\kappa_{n,\gamma}$ and $\kappa_n$ interchangeably when unit tangent vectors are involved.

Using Lemma 9.9.5 and Theorem 1.8.10 one can show that $T_{\mathbf{p}}(\mathbf{S})$ has an orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2)$ so that

$$\begin{aligned} D\mathbf{n}(\mathbf{p})(\mathbf{u}_1) &= -\kappa_1\mathbf{u}_1 \\ D\mathbf{n}(\mathbf{p})(\mathbf{u}_2) &= -\kappa_2\mathbf{u}_2. \end{aligned} \tag{9.45}$$

Furthermore, if we assume that $\kappa_1 \geq \kappa_2$, then $\kappa_1$ and $\kappa_2$ are the maximum and minimum of the second fundamental form $Q_{II}$, respectively, when restricted to the unit circle of the tangent space at $\mathbf{p}$. The numbers $\kappa_1$ and $\kappa_2$ are also the maximum and minimum, respectively, of the values of the normal curvatures $\kappa_n$ at $\mathbf{p}$.

**Definition.** The numbers $\kappa_1$ and $\kappa_2$ are called the *principal normal curvatures* to $\mathbf{S}$ at $\mathbf{p}$. The unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are called the *principal normal directions* for $\mathbf{S}$ at $\mathbf{p}$.

The principal normal directions will be unique at those places where $D\mathbf{n}$ is nonsingular. These definitions agree with those that were given at the beginning of this section. Furthermore, we can now easily prove Euler's formula (9.41) for the normal curvature in any direction. Let $\mathbf{u}$ be a unit vector in $T_{\mathbf{p}}(\mathbf{S})$. Then using the orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2)$ used in Equation (9.45) we can express $\mathbf{u}$ in the form

$$\mathbf{u} = (\cos\theta)\mathbf{u}_1 + (\sin\theta)\mathbf{u}_2$$

for some angle $\theta$. It follows that

$$\begin{aligned} \kappa_n &= Q_{II}(\mathbf{u}) \\ &= -D\mathbf{n}(\mathbf{p})(\mathbf{u}) \bullet \mathbf{u} \\ &= -D\mathbf{n}(\mathbf{p})((\cos\theta)\mathbf{u}_1 + (\sin\theta)\mathbf{u}_2) \bullet ((\cos\theta)\mathbf{u}_1 + (\sin\theta)\mathbf{u}_2) \\ &= ((\kappa_1\cos\theta)\mathbf{u}_1 + (\kappa_2\sin\theta)\mathbf{u}_2) \bullet ((\cos\theta)\mathbf{u}_1 + (\sin\theta)\mathbf{u}_2) \\ &= \kappa_1\cos^2\theta + \kappa_2\sin^2\theta. \end{aligned}$$

**Definition.** Let $\mathbf{S}$ be a surface and let $D\mathbf{n}(\mathbf{p}): T_{\mathbf{p}}(\mathbf{S}) \to T_{\mathbf{p}}(\mathbf{S})$ be the Weingarten map at a point $\mathbf{p}$ in $\mathbf{S}$. The determinant of $D\mathbf{n}(\mathbf{p})$ is called the *Gauss curvature* of $\mathbf{S}$ at $\mathbf{p}$ and is denoted by $K(\mathbf{p})$. The *mean curvature* of $\mathbf{S}$ at $\mathbf{p}$, denoted by $H(\mathbf{p})$, is derived from the trace of $D\mathbf{n}(\mathbf{p})$ and defined by

$$H(\mathbf{p}) = -(1/2)\ \mathrm{tr}(D\mathbf{n}(\mathbf{p})).$$

**9.9.8. Theorem.** The two definitions of Gauss curvature agree.

**Proof.** See [DoCa76] or [Spiv70b].

**9.9.9. Theorem.**

(1) $K(\mathbf{p}) = \boldsymbol{\kappa}_1 \boldsymbol{\kappa}_2$.
(2) $H(\mathbf{p}) = (\boldsymbol{\kappa}_1 + \boldsymbol{\kappa}_2)/2$.

**Proof.** Part (1) follows from equations (9.45) and the fact that the determinant of a linear transformation is independent of a basis. Part (2) follows from a similar property for the trace function.

To get a little feel for this new definition of Gauss curvature, we rework Examples 9.9.1–9.9.3.

**9.9.10. Example.** Let **S** be a sphere of radius r about a point **c**. Then the Gauss map is defined by

$$\mathbf{n}(\mathbf{p}) = \frac{\mathbf{p} - \mathbf{c}}{|\mathbf{p} - \mathbf{c}|} = \frac{1}{r}(\mathbf{p} - \mathbf{c})$$

and

$$D\mathbf{n}(\mathbf{p})(\gamma'(0)) = (\mathbf{n} \circ \gamma)'(0) = \frac{1}{r}\gamma'(0).$$

Therefore, a matrix for $D\mathbf{n}(\mathbf{p})$ is

$$\begin{pmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{r} \end{pmatrix},$$

so that $K = 1/r^2$ and $H = -1/r$. We can also see that $\boldsymbol{\kappa}_1 = \boldsymbol{\kappa}_2 = -1/r$.

**9.9.11. Example.** If **S** is a plane, then the Gauss map is constant, so that $D\mathbf{n}(\mathbf{p}) = 0$ and the matrix for $D\mathbf{n}(\mathbf{p})$ is zero. It follows that $K = H = 0$.

**9.9.12. Example.** Let S be the cylinder of radius r about the z-axis defined by

$$x^2 + y^2 = r^2.$$

The Gauss map is defined by

$$\mathbf{n}(\mathbf{p}) = \frac{1}{r}(\mathbf{p} - \mathbf{p} \bullet \mathbf{e}_3)$$

and

$$D\mathbf{n}(\mathbf{p})(\gamma'(0)) = (\mathbf{n} \circ \gamma)'(0) = \frac{1}{r}(\gamma'(0) - \mathbf{e}_3 \bullet \gamma'(0)).$$

We would like to find an orthonormal basis $(\mathbf{u}_1,\mathbf{u}_2)$ of eigenvectors of $D\mathbf{n}(\mathbf{p})$. If

$$\mathbf{p} = (r\cos\theta_0, r\sin\theta_0, p_3),$$

then let $\mathbf{u}_1 = (-\sin\theta_0, \cos\theta_0, 0)$ and let $\mathbf{u}_2 = \mathbf{e}_3$. Note that the curve

$$\gamma(\theta) = \left(r\cos\left(\frac{\theta}{r} + \theta_0\right), r\sin\left(\frac{\theta}{r} + \theta_0\right), p_3\right)$$

has the property that $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{u}_1$. It is now easy to check from the formula for $D\mathbf{n}(\mathbf{p})$ above that

$$D\mathbf{n}(\mathbf{p})(\mathbf{u}_1) = (1/r)\mathbf{u}_1$$
$$D\mathbf{n}(\mathbf{p})(\mathbf{u}_2) = 0.$$

This shows that $K = 0$ and $H = -1/2r$.

Although we do not have time to discuss mean curvature here, it certainly has geometrical significance. Here are two facts:

(1) The mean curvature at a point is the average of the normal curvatures there. See [MilP77] or [DoCa76].
(2) The mean curvature plays an important role in the study of surfaces that have minimum area for a fixed boundary.

**Plateau's Problem:** Given a closed curve $\mathbf{C}$, to find the surface of minimum area that has $\mathbf{C}$ for its boundary.

Plateau's problem is an old one and solving it has motivated a lot of research over the years. It is well known that soap films spanning wireframe boundaries create minimum area surfaces. It can be shown that a necessary condition for a surface to minimizes area is that its mean curvature vanish. Therefore another active research area is the study of minimal surfaces.

**Definition.** A *minimal surface* is a surface for which the mean curvature vanishes everywhere.

Minimal surfaces arise in other contexts, not just where one is minimizing area. See [Osse69] or [Gray98]. There are lots of questions here about their existence, uniqueness, construction, and characterization.

**Definition.** The *Dupin indicatrix* at a point $\mathbf{p}$ of the surface $\mathbf{S}$ is the subset $\mathbf{X}$ of $T_{\mathbf{p}}(\mathbf{S})$ defined by

$$\mathbf{X} = \{\mathbf{u} \in T_{\mathbf{p}}(\mathbf{S}) \mid Q_{II}(\mathbf{u}) = \pm 1\}.$$

**9.9.13. Theorem.** Let $\mathbf{p}$ be a point on a surface $\mathbf{S}$ with Gauss curvature K. Then

(1) If $K > 0$, then the Dupin indicatrix is an ellipse.
(2) If $K < 0$, then the Dupin indicatrix consists of two pairs of hyperbolas.

(3) If $K = 0$, then the Dupin indicatrix consists of either two parallel lines if one principal curvature is not zero or is empty if both principal curvatures are zero.

**Proof.** This is an immediate consequence of Euler's theorem and Theorem 9.9.9(1).

Theorem 9.9.13 motivates the following definition of an elliptic, hyperbolic, or parabolic point on a surface. Another common definition of these terms can be found in Exercise 9.9.6.

**Definition.** Let $\mathbf{p}$ be a point of a surface $\mathbf{S}$ and let K be the Gauss curvature at $\mathbf{p}$. The point $\mathbf{p}$ is called

(1) *elliptic* if $K > 0$,
(2) *hyperbolic* if $K < 0$,
(3) *parabolic* if $K = 0$ but $D\mathbf{n}(\mathbf{p}) \neq 0$ (only one principal curvature vanishes), and
(4) *planar* or *flat* if $D\mathbf{n}(\mathbf{p}) = 0$ (both principal curvatures vanish).

Using Theorem 9.9.9 one can make the following observations. See Figure 9.19 again. At an elliptic point both principal curvatures have the same sign. This means that all curves through that point must have their normal vector point to the same side of tangent plane. Spheres and ellipsoids are examples of such surfaces (see Figure 9.19(a)). At a hyperbolic point $\mathbf{p}$, the principal curvatures have opposite signs and so one can find two curves through $\mathbf{p}$ whose normals at $\mathbf{p}$ point to opposite sides of the tangent plane. A saddle surface, such as the hyperbolic paraboloid

$$z = x^2 - y^2$$

(see Figure 9.19(b)) and the point (0,0,0), is an example. The cylinder (Figure 9.19(c)) is an example of a parabolic point. One of the principal curvatures is zero, but the other is not. Points of a plane are example of planar points, but nonplanar surfaces can have such points. For example, (0,0,0) is a planar point of the surface of revolution where the curve $z = x^4$ is revolved about the z-axis (Exercise 9.9.1). Finally, we should mention the classical interpretation of the Dupin indicatrix, namely, if we take the tangent plane at a point $\mathbf{p}$ and intersect it with the surface $\mathbf{S}$ after moving it slightly in the normal direction, then the intersection curve will be the same sort of curve as the Dupin indicatrix. Figure 9.23(a) and (b) shows the intersection with the moved plane $\mathbf{X}$ at an elliptic and parabolic point $\mathbf{p}$, respectively. See [DoCa76] or [MilP77].

**Definition.** A point on a surface where the principal normal curvatures are equal ($\kappa_1 = \kappa_2$) is called an *umbilical point*.

Planar points are umbilical points.

**9.9.14. Theorem.** If every point of a connected surface $\mathbf{S}$ is an umbilical points, then $\mathbf{S}$ is either contained in a sphere or in a plane.

**Proof.** See [DoCa76].


**Figure 9.23.** The classical interpretation of the Dupin indicatrix.

**Definition.** An *asymptotic direction* at a point $\mathbf{p}$ of a surface $\mathbf{S}$ is a nonzero vector $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{S})$ so that $Q_{II}(\mathbf{v}) = 0$. A curve in $\mathbf{S}$ with the property that its tangent vector at every point is an asymptotic direction at that point is called an *asymptotic line*.

An elliptic point has no asymptotic directions. At a hyperbolic point there are two linearly independent asymptotic directions corresponding to two lines through the origin of the tangent plane. This matches the terminology with regard to ordinary hyperbolas in the plane. Parabolic points have a single line of asymptotic directions. The silhouette line along the top of the surface in Figure 9.23(b) defines an asymptotic line.

Next, we want to express the second fundamental form in local coordinates similar to what we did for the first fundamental form. Let $\mathbf{U}$ be an open set in the uv-plane and let

$$\Phi : \mathbf{U} \to \mathbf{V} \subseteq \mathbf{S}$$

be a regular parameterization of a neighborhood $\mathbf{V}$ of a point $\mathbf{p}$ in the surface $\mathbf{S}$ in $\mathbf{R}^3$. Let $\mathbf{n}(\mathbf{p})$ denote the unit normal vector to $\mathbf{S}$ as before. Assume $\mathbf{U}$ contains the origin and that $\Phi(\mathbf{0}) = \mathbf{p}$. Let $\mathbf{v}$ be a tangent vector in the tangent space of $\mathbf{S}$ at $\mathbf{p}$. Choose a curve

$$\gamma : (-a,a) \to \mathbf{V}$$

in $\mathbf{S}$ with $\gamma(\mathbf{0}) = \mathbf{p}$ and $\mathbf{v} = \gamma'(0)$. Express $\gamma$ in the form $\gamma(t) = \Phi(\mu(t))$, where

$$\mu : (-a,a) \to \mathbf{R}^2$$

and $\mu(0) = \mathbf{0}$. See Figure 9.16 again. Let $\mu(t) = (u(t),v(t))$.

To simplify the notation, we define

$$\Phi_u = \frac{\partial \Phi}{\partial u}, \quad \Phi_v = \frac{\partial \Phi}{\partial v}, \quad \Phi_{uu} = \frac{\partial^2 \Phi}{\partial u^2}, \quad \text{etc.}$$

Further, we abbreviate $\mathbf{n}(\Phi(u,v))$ to $\mathbf{n}(u,v)$ and define

$$\mathbf{n}_u(u,v) = \frac{\partial \mathbf{n}}{\partial u}(u,v) \equiv \frac{\partial(\mathbf{n}\circ\Phi)}{\partial u}(u,v) \quad \text{and} \quad \mathbf{n}_v(u,v) = \frac{\partial \mathbf{n}}{\partial v}(u,v) \equiv \frac{\partial(\mathbf{n}\circ\Phi)}{\partial v}(u,v).$$

The chain rule implies that

$$D\mathbf{n}(\gamma(t))(\gamma'(t)) = \mathbf{n}_u(u,v)u'(t) + \mathbf{n}_v(u,v)v'(t). \tag{9.46}$$

It follows from equations (9.46) and (9.35) that

$$\begin{aligned} Q_{II}(\gamma') &= -D\mathbf{n}(\mathbf{p})(\gamma') \bullet \gamma' \\ &= -(\mathbf{n}_u u' + \mathbf{n}_v v') \bullet (\Phi_u u' + \Phi_v v') \\ &= Lu'^2 + 2Mu'v' + Nv'^2, \end{aligned} \tag{9.47}$$

where, using the identities that one gets when the known equations $\mathbf{n} \bullet \Phi_u = 0 = \mathbf{n} \bullet \Phi_v$ are differentiated,

$$\begin{aligned} L &= -\mathbf{n}_u \bullet \Phi_u = \mathbf{n} \bullet \Phi_{uu}, \\ M &= -\mathbf{n}_v \bullet \Phi_u = \mathbf{n} \bullet \Phi_{uv} = \mathbf{n} \bullet \Phi_{vu} = -\mathbf{n}_u \bullet \Phi_v, \quad \text{and} \\ N &= -\mathbf{n}_v \bullet \Phi_v = \mathbf{n} \bullet \Phi_{vv}. \end{aligned}$$

**Definition.** The functions L, M, and N are called the *coefficients of the second fundamental form*.

Note that the matrix

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

is just the matrix of the symmetric bilinear map associated to the quadratic form $Q_{II}$ with respect to the basis consisting of the vectors $\Phi_u$ and $\Phi_v$ and $LN - M^2$ is the discriminant of $Q_{II}$ with respect to that basis. In terms of the shape operator S we have

$$\begin{aligned} L &= S(\Phi_u) \bullet \Phi_u, \\ M &= S(\Phi_v) \bullet \Phi_u = S(\Phi_u) \bullet \Phi_v, \quad \text{and} \\ N &= S(\Phi_v) \bullet \Phi_v. \end{aligned}$$

Note also that equation (9.47) expresses the Dupin indicatrix in terms of the functions L, M, and N. Asymptotic lines are defined by the equation

$$L(u')^2 + 2M(u')(v') + N(v')^2 = 0.$$

Next, compare Equation (9.47) for the second fundamental form with Equation (9.39) for the first fundamental form. The functions L, M, and N play just as impor-

tant a role as the functions E, F, and G. In particular, let us show that the derivative of the Gauss map can be described in terms of these functions. First of all, since $\mathbf{n}_u(u,v)$ and $\mathbf{n}_v(u,v)$ lie in the tangent space at $\Phi(u,v)$ ($\mathbf{n}\bullet\mathbf{n} = 1$ implies $\mathbf{n}_u\bullet\mathbf{n} = \mathbf{n}_v\bullet\mathbf{n} = 0$), we can express the vectors in terms of the local basis, namely,

$$\begin{aligned}\mathbf{n}_u &= c_{11}\Phi_u + c_{12}\Phi_v\\ \mathbf{n}_v &= c_{21}\Phi_u + c_{22}\Phi_v\end{aligned} \tag{9.48}$$

for some constants $c_{ij}$. Equations (9.46) and (9.48) imply that

$$D\mathbf{n}(\gamma') = (c_{11}u' + c_{21}v')\Phi_u + (c_{12}u' + c_{22}v')\Phi_v \tag{9.49}$$

and that the matrix for $D\mathbf{n}$ with respect to the basis $(\Phi_u,\Phi_v)$ is

$$\begin{pmatrix} c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}. \tag{9.50}$$

But from equation (9.48) and the definition of the coefficients of the first fundamental form we get

$$\begin{aligned}-L &= \mathbf{n}_u\bullet\Phi_u = c_{11}E + c_{12}F\\ -M &= \mathbf{n}_u\bullet\Phi_v = c_{11}F + c_{12}G\\ -M &= \mathbf{n}_v\bullet\Phi_u = c_{21}E + c_{22}F\\ -N &= \mathbf{n}_v\bullet\Phi_v = c_{21}F + c_{22}G.\end{aligned} \tag{9.51}$$

We can express equations (9.51) in matrix form as

$$-\begin{pmatrix} L & M\\ M & N\end{pmatrix} = \begin{pmatrix} c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}\begin{pmatrix} E & F\\ F & G\end{pmatrix}. \tag{9.52}$$

This is easily solved for the matrix $(c_{ij})$ to establish that

$$c_{11} = \frac{MF - LG}{EG - F^2}, \tag{9.53a}$$

$$c_{12} = \frac{LF - ME}{EG - F^2}, \tag{9.53b}$$

$$c_{21} = \frac{NF - MG}{EG - F^2}, \tag{9.53c}$$

$$c_{22} = \frac{MF - NE}{EG - F^2}. \tag{9.53d}$$

Equations (9.53) complete our analysis of D**n** and we are ready to draw some important consequences. Incidentally, equations (9.48) with the values of $c_{ij}$ as shown in (9.53) are called the *Weingarten equations*.

**9.9.15. Theorem.** The Gauss curvature K satisfies

$$K = \frac{LN - M^2}{EG - F^2} = \frac{LN - M^2}{|\Phi_u \times \Phi_v|^2}.$$

**Proof.** The theorem follows from equation (9.52) and the definition of Gauss curvature as the determinant of matrix (9.50). We see that K is just the quotient of the discriminants of the first and second fundamental forms.

**9.9.16. Theorem.** The mean curvature H satisfies

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)}.$$

**Proof.** The theorem follows from the definition of H and the values in (9.53) from which we can immediately compute the trace of the matrix (9.50) for D**n**.

Finally, we relate the Gauss and mean curvatures to the principal normal curvatures $\kappa_1$ and $\kappa_2$. Since the $-\kappa_1$ and $-\kappa_2$ are eigenvalues of D**n**, the linear map D**n** + kI is not invertible when $k = \kappa_1$ or $\kappa_2$. Therefore, its matrix has zero determinant, that is,

$$\det\begin{pmatrix} c_{11} + k & c_{12} \\ c_{21} & c_{22} + k \end{pmatrix} = 0.$$

This expands to

$$k^2 + (c_{11} + c_{22})k + (c_{11}c_{22} - c_{12}c_{21}) = 0,$$

or

$$k^2 - 2Hk + K = 0, \tag{9.54}$$

for $k = \kappa_1$ or $\kappa_2$. In other words, we have proved

**9.9.17. Theorem.** The principal normal curvatures satisfy Equation (9.54) and are given by the formula

$$H \pm \sqrt{H^2 - K}. \tag{9.55}$$

It is again time for an example. Analyzing the geometry invariably involves computing the functions that are the coefficients of the first and second fundamental forms. The functions L, M, and N are the more complicated. They were defined in the derivation of equation (9.47). Each had basically two definitions. However, given the

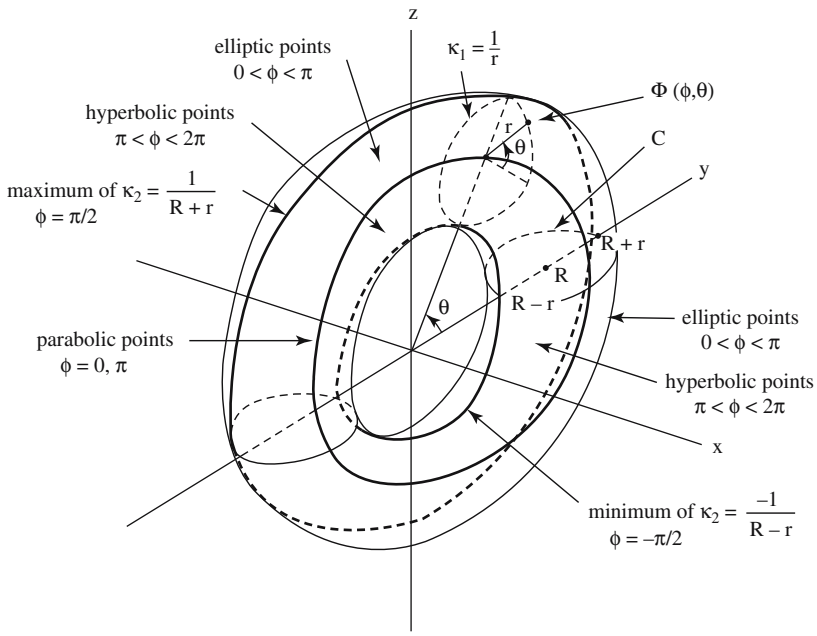



**Figure 9.24.** The torus.

fact that the formula for the derivative of the function $\mathbf{n}(\Phi(u,v))$ may be complicated, one usually prefers to use the following formulas:

$$L = \Phi_{uu} \bullet \mathbf{n}, \quad M = \Phi_{uv} \bullet \mathbf{n}, \quad \text{and} \quad N = \Phi_{vv} \bullet \mathbf{n}. \tag{9.56}$$

**9.9.18. Example.** Consider the torus that is obtained by rotating the circle of radius r with center (0,R,0), r < R, in the xy-plane about the x-axis. See Figure 9.24. It can be parameterized by

$$\Phi(\phi,\theta) = (r\cos\phi, (R + r\sin\phi)\cos\theta, (R + r\sin\phi)\sin\theta).$$

One can check that

$$\begin{aligned}
\Phi_\phi(\phi,\theta) &= (-r\sin\phi, r\cos\phi\cos\theta, r\cos\phi\sin\theta)\\
\Phi_\theta(\phi,\theta) &= (0, -(R + r\sin\phi)\sin\theta, (R + r\sin\phi)\cos\theta)\\
\mathbf{n}(\phi,\theta) &= \frac{\Phi_\phi(\phi,\theta) \times \Phi_\theta(\phi,\theta)}{|\Phi_\phi(\phi,\theta) \times \Phi_\theta(\phi,\theta)|} = (\cos\phi, \sin\phi\cos\theta, \sin\phi\sin\theta)\\
\Phi_{\phi\phi}(\phi,\theta) &= (-r\cos\phi, -r\sin\phi\cos\theta, -r\sin\phi\sin\theta)\\
\Phi_{\phi\theta}(\phi,\theta) &= (0, -r\cos\phi\sin\theta, r\cos\phi\cos\theta)\\
\Phi_{\theta\theta}(\phi,\theta) &= (0, -(R + r\sin\phi)\cos\theta, -(R + r\sin\phi)\sin\theta).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E &= \Phi_\phi \bullet \Phi_\phi = r^2 \\
F &= \Phi_\phi \bullet \Phi_\theta = 0 \\
G &= \Phi_\theta \bullet \Phi_\theta = (R + r\sin\phi)^2 \\
EG - F^2 &= r^2(R + r\sin\phi)^2 \qquad \left(= |\Phi_\phi \times \Phi_\theta|^2\right) \\
L &= \Phi_{\phi\phi} \bullet \mathbf{n} = -r \\
M &= \Phi_{\phi\theta} \bullet \mathbf{n} = 0 \\
M &= \Phi_{\theta\theta} \bullet \mathbf{n} = -(R + r\sin\phi)\sin\phi \\
LN - M^2 &= r(R + r\sin\phi)\sin\phi.
\end{aligned}$$

These formulas and the preceding Theorems allow us to conclude that

$$K = \frac{LN - M^2}{EG - F^2} = \frac{\sin\phi}{r(R + r\sin\phi)}$$

and

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)} = -\frac{1}{2}\left(\frac{\sin\phi}{R + r\sin\phi} + \frac{1}{r}\right).$$

Since the sign of K is determined by $LN - M^2$, we get elliptic points when $0 < \phi < \pi$, hyperbolic points when $\pi < \phi < 2\pi$, and parabolic points ($K = 0$) when $\phi = 0$ or $\pi$. Next, Equation (9.24) implies that

$$\kappa_1 = \frac{1}{r} \quad \text{and} \quad \kappa_2 = \frac{\sin\phi}{r(R + r\sin\phi)}.$$

The maximum normal curvature $\kappa_1$ is the same everywhere and is the curvature of the circle that generates the torus. The minimum normal curvature $\kappa_2$ varies with the meridians specified by $\phi$. It has its maximum on the outside where $\phi = \pi/2$, its minimum on the inside where $\phi = -\pi/2$, and is zero when $\phi = 0$ or $\pi$.

We pause for a moment to investigate the degree to which what we have done so far depends on the choice of local coordinates. The definition of the second fundamental form is certainly independent of the coordinates. On the other hand, had we chosen to base the definition on equation (9.47), as is done sometimes, there would have been a question. Exercise 9.9.5 asks the reader to show the independence of local coordinates of the definition in (9.47). The proof involves the chain rule similar to what we did in showing the independence of equation (9.37) for the first fundamental form. The coefficients of the second fundamental form do depend on the coordinate system but behave like the coefficients of the first fundamental form.

**9.9.19. Theorem.** In the case of surfaces the coefficients of the second fundamental form transform as follows (using the notation in Theorem 9.8.12):

$$L(\Phi) = L(\psi)u_x^2 + 2M(\psi)u_x v_x + N(\psi)v_x^2$$
$$M(\Phi) = L(\psi)u_x u_y + M(\psi)(u_x v_y + v_x u_y) + N(\psi)v_x v_y$$
$$N(\Phi) = L(\psi)u_y^2 + 2M(\psi)u_y v_y + N(\psi)v_y^2$$

**Proof.** Use the appropriate chain rule.

We have not paid much attention to the local parameterizations that we used. By choosing these more carefully we can make formulas simpler (as in the case of curves where arc-length parameterization was superior to other ones).

**Definition.** A *line of curvature* on a surface is a curve whose tangent at every point is parallel to a principal direction.

**9.9.20. Theorem.** (Rodrigues) Let $\Phi(u,v)$ be any regular parameterization of a surface **S**. Then a curve $\gamma(t)$ in **S** defines a line of curvature if and only if

$$\mathbf{n}'(t) = -\boldsymbol{\kappa}(t)\gamma'(t), \tag{9.57}$$

where $\mathbf{n}(t) = \mathbf{n}(\gamma(t))$ and $\boldsymbol{\kappa}(t)$ is the principal curvature of the curve at $\gamma(t)$.

**Proof.** If $\gamma(t)$ is a line of curvature, then $\gamma'(t)$ must be an eigenvector of $D\mathbf{n}$.

Equation (9.57) is called the *Rodrigues formula*.

**9.9.21. Theorem.** Every nonumbilical point on a $C^3$ surface has a neighborhood on which there exist two orthogonal families of lines of curvature.

**Proof.** Proving this theorem amounts to solving the differential equations defined by (9.57). Standard results about differential equations imply the existence and uniqueness of local solutions.

Often the following weaker form of Theorem 9.9.21 is sufficient. It is much easier to prove.

**9.9.22. Theorem.** Given a point **p** on a $C^2$ surface we can always find a regular parameterization $\Phi(u,v)$ of a neighborhood of **p** so that $\Phi_u$ and $\Phi_v$ at **p** are principal directions.

**Proof.** Choose any regular parameterization $\Phi(u,v)$ for a neighborhood of **p**. Assume that $\Phi(0,0) = \mathbf{p}$. If $\Phi_u$ and $\Phi_v$ are not already principal directions, we shall simply "rotate" the coordinate patch to achieve this. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be linearly independent vectors in $\mathbf{R}^2$ with the property that $D\Phi(\mathbf{u}_i)$ are principal directions and let T be the linear transformation of $\mathbf{R}^2$ so that $T(\mathbf{e}_i) = \mathbf{u}_i$. Then $\varphi = \Phi^{-1}T$ is the desired parameterization.

**9.9.23. Theorem.** Given a regular parameterization $\Phi(u,v)$ of a surface, the vectors $\Phi_u$ and $\Phi_v$ at a nonumbilical point are in the direction of the principal directions if and only if $F = M = 0$ at that point.

**Proof.** At a nonumbilical point the principal directions are orthogonal and by definition F = 0 if and only if $\Phi_u$ and $\Phi_v$ are orthogonal. If F = 0, then M = 0 if and only if the matrix in (9.50) is diagonal, which is equivalent to $\Phi_u$ and $\Phi_v$ being principal directions and Rodrigues's formula holding.

From this we get immediately

**9.9.24. Corollary.** Given a regular parameterization $\Phi(u,v)$ of a surface with no umbilical points, then the u- and v-parameter curves on the patch are lines of curvature if and only if F = M = 0.

**9.9.25. Theorem.** If the vectors $\Phi_u$ and $\Phi_v$ of a regular parameterization $\Phi(u,v)$ of a surface are the principal directions at a point, then the principal normal curvatures at that point are given by

$$\kappa_1 = \frac{L}{E} \quad \text{and} \quad \kappa_2 = \frac{N}{G}.$$

**Proof.** Given the hypothesis, Theorem 9.9.23 implies that F = M = 0. Substituting into the formulas for the Gauss curvature and mean curvature given in Theorems 9.9.15 and 9.9.16 and solving Equation (9.55) for this special case gives the desired two roots.

**9.9.26. Corollary.** If the u- and v-parameter curves of a regular parameterization $\Phi(u,v)$ of a surface are lines of curvature, then the principal curvatures at that point are given by

$$\kappa_1 = \frac{L}{E} \quad \text{and} \quad \kappa_2 = \frac{N}{G}.$$

We are almost done with our foray into the geometry of surfaces, but it seems appropriate to take it a little further and finish with the so-called fundamental theorem of surfaces (Theorem 9.9.28 below) that is the analog of Theorem 9.4.7 for curves. Even though most of the results in the rest of this section are stated without proofs (they are actually not that hard), the reader will at least get an overview of what needs to be done. For details see [DoCa76], [Lips69], [Spiv70b], or [Spiv75].

The main missing ingredient in what we have done so far is to express the second derivatives $\Phi_{uu}$, $\Phi_{uv}$, and $\Phi_{vv}$ in terms of the first derivatives and **n**. This can be done because the three vectors form a basis for $\mathbf{R}^3$. Suppose we write

$$\begin{aligned}
\Phi_{uu} &= \Gamma_{11}^1\Phi_u + \Gamma_{11}^2\Phi_v + \alpha_{11}\mathbf{n}\\
\Phi_{uv} &= \Gamma_{12}^1\Phi_u + \Gamma_{12}^2\Phi_v + \alpha_{12}\mathbf{n}\\
\Phi_{vu} &= \Gamma_{21}^1\Phi_u + \Gamma_{21}^2\Phi_v + \alpha_{21}\mathbf{n}\\
\Phi_{vv} &= \Gamma_{22}^1\Phi_u + \Gamma_{22}^2\Phi_v + \alpha_{22}\mathbf{n}
\end{aligned} \tag{9.58}$$

where $\Gamma_{ij}^{s}$ and $\alpha_{ij}$ are some constants. One can show (by taking the dot product of equations (9.28) with $\Phi_u$ and $\Phi_v$) that these constants are given by the following formulas:

$$\alpha_{11} = L, \quad \alpha_{12} = \alpha_{21} = M, \quad \alpha_{22} = N,$$

$$\Gamma_{11}^{1} = \frac{GE_u - 2FF_u + FE_v}{2(EG - F^2)}, \quad \Gamma_{12}^{1} = \Gamma_{21}^{1} = \frac{GE_v - FG_u}{2(EG - F^2)}, \quad \Gamma_{22}^{1} = \frac{2GF_v - GG_u - FG_v}{2(EG - F^2)},$$

$$\Gamma_{11}^{2} = \frac{2EF_u - EE_v + FE_u}{2(EG - F^2)}, \quad \Gamma_{12}^{2} = \Gamma_{21}^{2} = \frac{EG_u - FE_v}{2(EG - F^2)}, \quad \Gamma_{22}^{2} = \frac{EG_v - 2FF_v + FG_u}{2(EG - F^2)}. \tag{9.59}$$

**Definition.** The quantities $\Gamma_{ij}^{s}$ are called the *Christoffel symbols* of the surface **S** for the parameterization $\Phi(u,v)$.

Not only do we have the formulas above for the Christoffel symbols, but since they are functions of E, F, and G and their derivatives, it follows that **any quantity defined in terms of these symbols will be invariant under isometries**!

We get some additional constraints on the Christoffel symbols if we use the fact that $C^3$ functions have mixed partials that can be obtained by differentiating in any order. In particular,

$$(\Phi_u)_{uv} = (\Phi_u)_{vu} \quad \text{and} \quad (\Phi_v)_{uv} = (\Phi_v)_{vu}. \tag{9.60}$$

One can show that equations (9.60) hold if and only if

$$\begin{aligned} L_v - M_u &= L\,\Gamma_{12}^{1} + M(\Gamma_{12}^{2} - \Gamma_{11}^{1}) - N\,\Gamma_{11}^{2} \\ M_v - L_u &= L\,\Gamma_{22}^{1} + M(\Gamma_{22}^{2} - \Gamma_{12}^{1}) - N\,\Gamma_{12}^{2} \end{aligned} \tag{9.61}$$

and

$$\begin{aligned} LN - M^2 = F\,&\left[(\Gamma_{22}^{2})_u - (\Gamma_{12}^{2})_v + \Gamma_{22}^{1}\Gamma_{11}^{2} - \Gamma_{12}^{1}\Gamma_{12}^{2}\right] \\ &+ E\left[(\Gamma_{22}^{1})_u - (\Gamma_{12}^{1})_v + \Gamma_{22}^{1}\Gamma_{11}^{1} + \Gamma_{22}^{2}\Gamma_{12}^{1} - \Gamma_{12}^{1}\Gamma_{12}^{1} - \Gamma_{12}^{2}\Gamma_{22}^{1}\right]. \end{aligned} \tag{9.62}$$

Equations (9.61) are called the *Mainardi-Codazzi equations*. Equation (9.62) is called the *Gauss equation*. Equations (9.61) and (9.62) together are called the *compatibility equations*.

The Gauss equation has an important consequence. Since the Christoffel symbols $\Gamma_{ij}^{s}$ depend only on E, F, G, and their derivatives, the quantity $LN - M^2$ depends only on the coefficients of the first fundamental form and their derivatives. It follows from Theorem 9.9.15 that the Gauss curvature depends only on the coefficients of the first fundamental form (even though it was defined from the second fundamental form). Therefore we have proved one of the most fundamental results in the theory of surfaces, namely,

**9.9.27. Theorem.** (Theorema Egregrium) The Gauss curvature $K(\mathbf{p})$ of a surface of class $C^2$ is an isometric invariant.

Theorem 9.9.27 is the reason that mapmakers had so much trouble making a good map of the world. The converse of Theorem 9.9.27 is false in general, namely, a Gauss curvature-preserving map between surfaces is not necessarily an isometry. The converse **is** true in one special case, namely, the constant curvature case.

**9.9.28. Theorem.** Any two surfaces having the same constant Gauss curvature are isometric.

**Proof.** See [Stok69].

We finally come to our main result, namely, that knowledge of the first and second fundamental forms of a surface determine it completely locally.

**9.9.29. Theorem.** (The Fundamental Theorem of Surfaces) Let E, F, G, L, M, and N be differentiable functions defined on an open set $\mathbf{U}$ in $\mathbf{R}^2$ with E, G > 0. Assume

(1) $EG - F^2 > 0$ and
(2) E, F, G, L, M, and N satisfy the compatibility Equations (9.61) and (9.62).

Then, for each $\mathbf{p}$ in $\mathbf{U}$ there exists an open neighborhood $\mathbf{V}$ of $\mathbf{p}$ in $\mathbf{U}$, a surface $\mathbf{S}$ in $\mathbf{R}^3$, and a diffeomorphism

$$\Phi: \mathbf{V} \to \mathbf{S},$$

so that E, F, G, and L, M, N are the coefficients of the first and second fundamental form of $\mathbf{S}$, respectively. Furthermore, the surface $\mathbf{S}$ is unique up to rigid motion.

**Proof.** See [DoCa76], [Lips69], or [Spiv75]. The functions E, F, and G need only be $C^2$ and the functions L, M, and N, only $C^1$.

## 9.10 Geodesics

This section takes a look at how we might generalize the concept of a "straight" line that we have in $\mathbf{R}^n$. The generalization will be called a geodesic. We shall discuss geodesics in a surface in $\mathbf{R}^3$. The reason for concentrating on this special case is that it allows us to introduce the subject of geodesics in a geometrically intuitive way. The generalization to higher-dimensional manifolds with a Riemannian metric involves more advanced concepts from differential geometry that we shall only briefly get to in Section 9.17. We shall present four definitions for a geodesic. Although geodesics depend only on intrinsic properties of a surface, not all the definitions will have that property and will seem to rely on a surface's imbedding in $\mathbf{R}^3$. In practice this will not be a problem. In fact, it will facilitate certain computations.

Since a manifold is not a vector space, generalizing the simple definition of a straight line in terms of a point and direction vector will clearly not work. To generalize, one must understand what one is generalizing. What are some intuitive properties that one usually associates to a straight line? Here are several:

(1) Straight lines are not "curved", that is, they have zero curvature.
(2) Straight lines define paths of shortest lengths between two points.
(3) Between any two points in $\mathbf{R}^n$ there is a unique straight line from one to the other.
(4) The tangent vectors along the straight line are all parallel.

The different ways of looking at a straight line have a close connection to principles in physics, specifically, the mechanics of particles. It follows from basic principles of classical mechanics that a particle that is not under the influence of any external forces and that is moving in a surface will move along a geodesic in the surface. For example, the shortest path property reflects Jacobi's form of the principle of least action. The straightness of lines reflects the Hertz principle of least curvature. The shortest path property is probably one of the most common ways of thinking of a straight line. A nice elementary book that investigates shortest paths (and other variational problems) at great length is [Lyus64].

Before we get started generalizing the just-mentioned properties of straight lines, we want to clarify a point that might otherwise cause some confusion in the reader's mind later on. Basically, we again run into the problem that although we are really interested in characterizing certain **sets** (reading the just-mentioned properties certainly suggests that we are talking about paths or sets), the analysis will proceed by studying functions, namely, parameterizations of those sets. The term "geodesic" is applied to both paths, where there will be no confusion, and functions, where the confusion may arise with respect to the multiple definitions since there is no unique parameterization of a path. Of course, we could restrict ourselves to arc-length parameterizations. This is a good approach from a theoretical point of view, and will be our first one, because it would not only give us a unique parameterization for each path but also greatly simplify the analysis. The problem with **only** using arc-length parameterizations when dealing with geodesics is that one rarely sees these parameterizations in practice. Defining arc-length parameterizations is usually extremely complicated. For practical reasons therefore one want to study geodesics in the context of a class of functions that is at least broad enough to cover the kind of parameterizations one finds in the real world. The potential confusion to which we are referring here, and which is what we want to bring out into the open, arises from the fact that the classes of parameterizations, to which the different definitions for when a parameterization is a geodesic apply, are often different. This difference alone would therefore make the definitions different. For example, some definitions in the literature apply only to arc-length parameterizations, some to arbitrary regular parameterizations, and others to constant speed regular parameterizations. When using one of these definitions one must then of course be careful to apply it to a parameterization of the correct type. As long as the reader does that and understands what is going on, a question such as "is this definition the same as that one?" may be interesting from a technical point of view but is unimportant practically speaking. All definitions agree on arc-length parameterizations. They all lead to the **same** geodesic **paths**.

We start the discussion of geodesics by introducing a new aspect of the curvature of a space curve if it lies in a surface. Let

$$\Phi: \mathbf{U} \to \mathbf{S}$$

be a regular one-to-one parameterization of a surface $\mathbf{S}$ in $\mathbf{R}^3$ and let $\mathbf{n}(\mathbf{p})$ be the unit normal vector to the surface $\mathbf{S}$ at the point $\mathbf{p} = \Phi(u,v)$. Let $\gamma:[a,b] \to \mathbf{S}$ be a curve

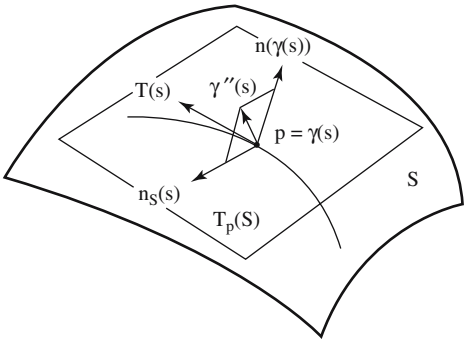



**Figure 9.25.** Defining geodesic curvature.

parameterized by **arc-length**. See Figure 9.25. As a curve in $\mathbf{R}^3$ we know that $\gamma(s)$ can be described in terms of its Frenet frame (T(s),N(s),B(s)), its curvature $\boldsymbol{\kappa}(s)$, and its torsion $\boldsymbol{\tau}(s)$. Define a unit vector $n_{\mathbf{S}}(s)$ so that $(T(s), n_{\mathbf{S}}(s), \mathbf{n}(\gamma(s)))$ is a frame that defines the standard orientation in $\mathbf{R}^3$. The vector $n_{\mathbf{S}}(s)$ will be tangent to **S** at $\gamma(s)$ and is defined by the equation

$$n_{\mathbf{S}}(s) = \mathbf{n}(\gamma(s)) \times T(s). \tag{9.63}$$

Recall that the second derivative $\gamma''(s)$ is closely related to the curvature of the **space** curve $\gamma(s)$. Since the tangent space to $\mathbf{R}^3$ at $\gamma(s)$ is the direct sum of the tangent space of **S** at $\gamma(s)$ and the one-dimensional orthogonal subspace with basis $\mathbf{n}(\gamma(s))$, we can write any vector, in particular $\gamma''(s)$, in the form

$$\gamma''(s) = W(s) + V(s) \tag{9.64}$$

where V(s) lies in the tangent space of **S** at $\gamma(s)$ and W(s) is a multiple of $\mathbf{n}(\gamma(s))$. The vector V(s) can also be described more explicitly. Since T(s) is tangent to **S**, $W(s) \bullet T(s) = 0$. But $\gamma(s)$ is arc-length parameterization, so that $\gamma''(s) \bullet T(s) = 0$. It follows from this and equation (9.64) that $V(s) \bullet T(s) = 0$, that is, V(s) is orthogonal to T(s). In other words, the vector V(s) is orthogonal to both $\mathbf{n}(\gamma(s))$ and T(s). Equation (9.63) now implies that V(s) is a multiple of $n_{\mathbf{S}}(s)$. Putting all these facts together implies that there are unique functions $\boldsymbol{\kappa}_n(s)$ and $\boldsymbol{\kappa}_g(s)$, so that

$$\gamma''(s) = \boldsymbol{\kappa}(s)N(s) = \boldsymbol{\kappa}_n(s)\mathbf{n}(\gamma(\mathbf{s})) + \boldsymbol{\kappa}_g(s)n_{\mathbf{S}}(s). \tag{9.65}$$

The value $\boldsymbol{\kappa}_n(s) = \gamma''(s) \bullet \mathbf{n}(\gamma(s))$ is of course just the normal curvature of the curve $\gamma(s)$ at s.

**Definition.** The value $\boldsymbol{\kappa}_g(s)$ is called the *geodesic curvature* of the **arc-length** parameterized curve $\gamma(s)$ at s.

The geodesic curvature function was introduced by F. Minding in 1830. Because we have orthonormal vectors, the following relation holds between $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}_n$, and $\boldsymbol{\kappa}_g$:

$$\boldsymbol{\kappa}^2 = \boldsymbol{\kappa}_n^2 + \boldsymbol{\kappa}_g^2. \tag{9.66}$$

Note also that, although $\boldsymbol{\kappa}$ is always nonnegative by definition, no such condition holds for $\boldsymbol{\kappa}_n$ or $\boldsymbol{\kappa}_g$. In fact, changing the direction of $\mathbf{n}(\gamma(s))$ changes the sign of both of these values. We already discussed the normal curvature of a curve in the last section. In this section it is the geodesic curvature that is interesting because that is what is needed to define geodesics.

Here are some facts that describe the geometry behind the function $\boldsymbol{\kappa}_g$ a little more.

**9.10.1. Theorem.** Let $\gamma(s)$ be a curve in a surface $\mathbf{S}$ in $\mathbf{R}^3$ parameterized by arc-length.

(1) If the surface $\mathbf{S}$ is a plane, then the geodesic curvature function $\boldsymbol{\kappa}_g(s)$ of the planar curve $\gamma(s)$ is just the ordinary **planar** signed curvature function of the curve $\gamma(s)$.
(2) $\boldsymbol{\kappa}_g(s) = \boldsymbol{\kappa}(s)\cos\alpha(s)$, where $\alpha(s)$ is the angle between the unit normal $\mathbf{n}(\gamma(s))$ and the binormal $B(s)$.
(3) $\boldsymbol{\kappa}_g(s)$ is the signed curvature at $\gamma(s)$ of the planar curve that is the orthogonal projection of $\gamma(s)$ onto the tangent plane of $\mathbf{S}$ at $\gamma(s)$.

**Proof.** In the case where $\mathbf{S}$ is the xy-plane, part (1) is an easy consequence of the definition of $n_{\mathbf{S}}(s)$ and the definition of curvature in a plane. Part (2) follows from the following string of equalities:

$$\begin{aligned}\boldsymbol{\kappa}_g(s) &= \gamma''(s)\bullet n_{\mathbf{S}}(s) = \gamma''(s)\bullet(\mathbf{n}(\gamma(s))\times T(s))\\ &= \boldsymbol{\kappa}(s)N(s)\bullet(\mathbf{n}(\gamma(s))\times T(s)) = \boldsymbol{\kappa}(s)\mathbf{n}(\gamma(s))\bullet(T(s)\times N(s))\\ &= \boldsymbol{\kappa}(s)\mathbf{n}(\gamma(s))\bullet B(s) = \boldsymbol{\kappa}(s)\cos\alpha(s).\end{aligned}$$

Part (3) generalizes part (1) and its truth follows intuitively from the formula in Part (2). The proof is left as Exercise 9.10.1.

We have considered the arc-length parameterization of curves in the discussion above because the formulas are simpler in that case; however, as usual, we are interested in geometric concepts associated to paths rather than their parameterizations. First, we extend the definitions to regular curves in the obvious way.

**Definition.** If $\eta(t)$ is a regular curve for which $\gamma(s) = \eta(\sigma(s))$ is the arc-length parameterization after an orientation-preserving change of parameters $t = \sigma(s)$, then the *normal curvature* $\boldsymbol{\kappa}_n(t)$ and the *geodesic curvature* $\boldsymbol{\kappa}_g(t)$ of $\eta(t)$ at t are defined to be the normal and geodesic curvatures, respectively, of the curve $\gamma(s)$ at the point $s = \sigma^{-1}(t)$.

Fortunately, we have a formula for computing the geodesic curvature for an arbitrary regular curve.

**9.10.2. Theorem.** If $\eta(t)$ be a regular curve in a surface $\mathbf{S}$ in $\mathbf{R}^3$, then

$$\boldsymbol{\kappa}_g(t) = \frac{\eta''(t)\bullet(\mathbf{n}(\eta(t))\times\eta'(t))}{|\eta'(t)|^3}.$$

**Proof.** Assume that $\eta(t) = \gamma(\alpha(t))$, where $\gamma(s)$ is the arc-length parameterization of the curve. Differentiating we get

$$\eta'(t) = \gamma'(\alpha(t))\alpha'(t) \quad \text{and} \quad \eta''(t) = \gamma''(\alpha(t))\alpha'^2(t) + \gamma'(\alpha(t))\alpha''(t).$$

Therefore, $|\eta'(t)| = |\alpha'(t)|$, (using the fact that $|\gamma'(s)| = 1$), and

$$\eta''(t) \times \eta'(t) = \alpha'^3(t)\gamma''(\alpha(t)) \times \gamma'(\alpha(t))$$

(using the fact that $\gamma' \times \gamma' = \mathbf{0}$). We now have the following string of equalities:

$$\begin{aligned}
\boldsymbol{\kappa}_g(t) &= \boldsymbol{\kappa}_g(\alpha(t)) = \gamma''(\alpha(t)) \bullet n_{\mathbf{S}}(\alpha(t)) \\
&= \gamma''(\alpha(t)) \bullet (\mathbf{n}(\gamma(\alpha(t))) \times \gamma'(\alpha(t))) \\
&= -\mathbf{n}(\gamma(\alpha(t))) \bullet (\gamma''(\alpha(t)) \times \gamma'(\alpha(t))) \\
&= -\frac{\mathbf{n}(\eta(t)) \bullet (\eta''(t) \times \eta'(t))}{|\alpha'(t)|^3} = \frac{\eta''(t) \bullet (\mathbf{n}(\eta(t)) \times \eta'(t))}{|\eta'(t)|^3}.
\end{aligned}$$

The theorem is proved.

We are ready to look at the problem of generalizing straight lines. We first generalize property (1) of a straight line, namely, that it is not curved. The tangent plane at a point of a surface is a good approximation to the surface. This suggests that we would like the projection of a geodesic onto the tangent plane to be a straight line, and so we take our cue from Theorem 9.10.1(3).

**First definition of a geodesic:** A *geodesic* in a surface $\mathbf{S}$ in $\mathbf{R}^3$ is a regular curve in $\mathbf{S}$ whose geodesic curvature is zero everywhere.

According to the definition, a geodesic is a **function**, but, as usual, we are really interested in properties of **paths**.

**Definition.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$. A subset $\mathbf{X}$ in $\mathbf{S}$ is called a *geodesic path*, or simply a *geodesic*, if there is a geodesic $\sigma:[a,b] \to \mathbf{S}$ with $\mathbf{X} = \sigma([a,b])$. A curve $\gamma:[c,d] \to \mathbf{S}$ is said to *generate a geodesic path* if $\gamma([c,d])$ is a geodesic path.

The next theorem contains some useful criteria for when regular curves are geodesics.

**9.10.3. Theorem.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$.

(1) If $\eta:[a,b] \to \mathbf{S}$ is a regular curve with the property that $\eta''(t)$ is orthogonal to $\mathbf{S}$ at $\eta(t)$ for all t, then $\eta(t)$ is a geodesic.
(2) A constant speed regular curve $\eta:[a,b] \to \mathbf{S}$ is a geodesic if and only if $\eta''(t)$ is orthogonal to $\mathbf{S}$ at $\eta(t)$ for all t.

**Proof.** If we have arc-length parameterization, then both parts of the theorem are obvious from Equation (9.65). Assume that $\eta(t)$ is a regular curve but not necessarily the arc-length parameterization.

If $\eta''(t)$ is orthogonal to **S**, then Theorem 9.10.2 easily implies that the geodesic curvature vanishes. This proves part (1) of the theorem and the "if" part for (2). To prove the "only if" part in (2), assume that $\eta(t)$ is a constant speed geodesic. Since the geodesic curvature of $\eta(t)$ vanishes, $\eta''(t)$ is a linear combination of the orthogonal vectors $\eta'(t)$ and $\mathbf{n}(\eta(t))$. This is all that we would be able to say in general, but $\eta(t)$ has constant speed, so that $\eta'(t) \bullet \eta'(t) = c$ for some constant c. Differentiating this equation shows that $\eta''(t)$ is orthogonal to $\eta'(t)$ and hence must be parallel to $\mathbf{n}(\eta(t))$, which is what we had to prove.

Note that the converse to Theorem 9.10.3(1) is false. Consider the (nonconstant speed) planar curve $\eta(t) = (t^2,t^2)$, $t \in [1,2]$. Since $\eta''(t) = (2,2)$, $\eta''(t)$ is not orthogonal to the plane even though its path, a straight line segment, is obviously a geodesic. This example shows that all we can say about $\eta''(t)$ for a regular curve $\eta(t)$ that is a geodesic is that $\eta''(t)$ is a linear combination of the vectors $\eta'(t)$ and $\mathbf{n}(\eta(t))$. Even so, because checking whether $\eta''$ is orthogonal to the surface is such a simple test, some texts turn the property into a definition.

**Second definition of a geodesic:** A *geodesic* in a surface **S** in $\mathbf{R}^3$ is a regular curve $\eta(t)$ in **S** with the property that $\eta''(t)$ is orthogonal to the surface at $\eta(t)$ for all t.

As we just saw, this is actually a stronger condition than necessary since it forces a geodesic to be a **constant speed** regular curve. It is therefore technically not equivalent to our first definition.

Because constant speed regular curves will appear frequently in this section, we collect two of their important properties in a theorem.

**9.10.4. Theorem.** Let $\eta(t)$ be a regular curve in a surface **S** in $\mathbf{R}^3$.

(1) If $\eta''(t)$ is orthogonal to the surface, then $\eta(t)$ is a constant speed curve.
(2) If $\eta(t)$ has constant speed, then the parameter t is proportional to the arc-length parameter s.

**Proof.** To prove (1), note that

$$(\eta' \bullet \eta')' = 2\,\eta' \bullet \eta'' = 0$$

implies that $\eta' \bullet \eta'$ is a constant function. To prove (2), assume that $|\eta'(t)|$ is equal to a constant $c \neq 0$, and check that $\gamma(s) = \eta(s/c)$ is the arc-length parameterization, that is, $s = ct$.

Here is a criterion that is more directed at the question of when a **path** is a geodesic by telling us what kind of parameterization we should seek.

**9.10.5. Theorem.** Let **S** be surface in $\mathbf{R}^3$. If $\eta:[a,b] \to \mathbf{S}$ is a regular curve with the property that the vector $\eta''(t)$ lies in the plane spanned by $\eta'(t)$ and $\mathbf{n}(\eta(t))$ for all t, then the arc-length parameterization of the curve $\eta(t)$ is a geodesic in **S**.

**Proof.** Let $\alpha:[0,L] \to [a,b]$ be the reparameterization of $\eta(t)$, so that $\gamma(s) = \eta(\alpha(s))$ is the arc-length parameterization. We have that

$$\gamma'(s) = \alpha'(s)\eta'(\alpha(s)) \quad \text{and} \quad \gamma''(s) = \alpha''(s)\eta'(\alpha(s)) + \alpha'(s)^2\eta''(\alpha(s)),$$

from which it follows that the vector $\gamma''(s)$ lies in plane spanned by $\eta'(\alpha(s))$ and $\eta''(\alpha(s))$. Our hypothesis about $\eta''$ now implies that the vector $\gamma''(s)$ lies in plane spanned by $\gamma'(s)$ and $\mathbf{n}(\gamma(s))$. But $\gamma'(s)$ and $\mathbf{n}(\gamma(s))$ are orthogonal vectors and so are $\gamma'(s)$ and $\gamma''(s)$ since $\gamma(s)$ is arc-length parameterization. It follows that $\gamma''(s)$ is a multiple of $\mathbf{n}(\gamma(s))$, that is, $\gamma''(s)$ is orthogonal to $\mathbf{S}$ at $\gamma(s)$. The theorem is proved.

**9.10.6. Example.** Lines in $\mathbf{R}^2$ are geodesics. Simply use a parameterization of the form $\gamma(t) = \mathbf{p}_0 + t\mathbf{v}$ for which $\gamma''(t) = 0$. More generally, a similar argument shows that if a surface $\mathbf{S}$ contains a segment, then that segment is a geodesic in $\mathbf{S}$.

**9.10.7. Example.** The curve

$$\gamma(t) = (\cos(at+b), \sin(at+b), ct+d), \tag{9.67}$$

where a, b, c, and d are constants, is a geodesic in the cylinder $x^2 + y^2 = 1$. Conversely, every geodesic of this cylinder has a parameterization of the form shown in Equation (9.67). See Figure 9.26.

**Proof.** By inspection

$$\gamma''(t) = \left(-a^2\cos(at+b), -a^2\sin(at+b), 0\right)$$

is orthogonal to the cylinder at $\gamma(t)$. To prove this directly, let

$$f(x,y,z) = x^2 + y^2 - 1.$$

The cylinder is just the zero set of the function $f(x,y,z)$ and we know that the gradient $\nabla f = (2x,2y,0)$ is a normal vector to the surface at $(x,y,z)$. This vector is parallel to $\gamma''(t)$.

To show the converse, let $\gamma(s)$ be an arc-length parameterized curve that is a geodesic and use equation (9.65) to deduce that

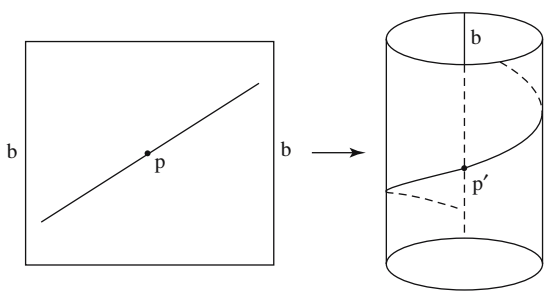



**Figure 9.26.** Geodesics on a cylinder.

$$T'(s) \bullet \mathbf{e}_3 = \gamma''(s) \bullet \mathbf{e}_3 = \boldsymbol{\kappa}_n(s)\mathbf{n}(\gamma(s)) \bullet \mathbf{e}_3 = 0.$$

Therefore, integration implies that $T(s) \bullet \mathbf{e}_3 = c$, for some constant c. Integrating again implies that $\gamma(s) \bullet \mathbf{e}_3 = cs + d$, for another constant d. It follows that

$$\gamma(s) = (\cos u(s), \sin u(s), cs + d)$$

for some function u(s). It is easy to check by differentiating that the condition $\gamma'(s) \bullet \gamma''(s) = 0$ implies that

$$u'(s)u''(s) = \left(\frac{1}{2}u'^2(s)\right)' = 0.$$

Two more integrations then imply that $u(s) = as + b$, for constants a and b. A more elegant way to solve the problem is to observe that

$$\Phi(u, v) = (\cos u, \sin u, v)$$

defines a local isometry of the plane with the cylinder. See Example 9.8.3 and Figure 9.26 again. Any geodesic on the cylinder must pull back to a straight line in the plane via this map since those are the only geodesics in the plane. We leave the reader to fill in the details.

**9.10.8. Example.** The *great circles*

$$\gamma(t) = (\cos at)\mathbf{u}_1 + (\sin at)\mathbf{u}_2, \tag{9.68}$$

where a and b are constants and $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal vectors, are geodesics in the sphere $\mathbf{S}^2$. Conversely, every geodesic of $\mathbf{S}^2$ has a parameterization of the form shown in equation (9.68).

**Proof.** To prove the first part, we again simply have to observe that

$$\gamma''(t) = \left(-a^2 \cos at\right)\mathbf{u}_1 + \left(-a^2 \sin at\right)\mathbf{u}_2 = -a^2\gamma(t)$$

is a vector orthogonal to the sphere at $\gamma(t)$.

To show the converse, let $\gamma(s)$ be the arc-length parameterization of a geodesic.

**Claim 1.** The curve $\gamma(s)$ has constant curvature. In fact, $\boldsymbol{\kappa} = 1$.

First of all, $\gamma' \bullet \gamma = 0$ implies that $\gamma'' \bullet \gamma + \gamma' \bullet \gamma' = \gamma'' \bullet \gamma + 1 = 0$, so that $\gamma'' \neq 0$ and $\gamma$ has a well-defined Frenet frame $(T(s), N(s), B(s))$. Since N(s) is also a unit normal to $\mathbf{S}^2$, the definition of the shape operator $S_{\mathbf{p}}$ for the sphere with respect to that normal vector field of $\mathbf{S}^2$ implies that

$$S_{\gamma(s)}(T(s)) = -DN(\gamma(s))(T(s)) = -N'(s) = \boldsymbol{\kappa}(s)T(s) - \boldsymbol{\tau}(s)B(s).$$

But we also know that

$$S_{\gamma(s)}(\mathbf{v}) = c\mathbf{v}$$

for all tangent vectors $\mathbf{v}$ of $\mathbf{S}^2$ at $\gamma(s)$ and some constant c because every point of the sphere $\mathbf{S}^2$ is an umbilical point. In our case of a unit sphere, c = 1. These two equations imply that $\boldsymbol{\kappa} = 1$ and $\boldsymbol{\tau} = 0$. Claim 1 is proved.

Let $\mathbf{p}_0 = \gamma(s_0)$ be a point on the curve.

**Claim 2.** The curve $\gamma(s)$ lies in the plane $\mathbf{X}$ through the origin with basis $\gamma'(s_0)$ and $\gamma''(s_0)$ $(= \mathbf{p}_0)$.

To prove Claim 2, note that $\gamma(s)$ is a unit normal to $\mathbf{S}^2$ at $\gamma(s)$, so that

$$(\gamma' \times \gamma'')' = (\gamma' \times \boldsymbol{\kappa}\gamma)' = \gamma'' \times \boldsymbol{\kappa}\gamma + \gamma' \times (\boldsymbol{\kappa}\gamma' + \boldsymbol{\kappa}'\gamma) = \mathbf{0}.$$

Therefore, $\gamma'(s) \times \gamma''(s)$ is a constant vector. If $\mathbf{b} = \gamma'(s_0) \times \gamma''(s_0)$, then

$$(\gamma(s) - \mathbf{p}_0) \bullet \mathbf{b} = 0,$$

which shows that $\gamma(s)$ lies in plane through $\mathbf{p}_0$ that has normal vector $\mathbf{b}$. Given the definition of $\mathbf{b}$, this proves Claim 2 and also shows that the curve lies in the great circle that is the intersection of the plane $\mathbf{X}$ and $\mathbf{S}^2$.

To finish the proof of the converse, pick an orthonormal basis $\mathbf{u}_1$ and $\mathbf{u}_2$ for the plane $\mathbf{X}$, express $\gamma(s)$ in the form

$$\gamma(s) = (\cos u(s))\mathbf{u}_1 + (\sin u(s))\mathbf{u}_2,$$

and show that u(s) has the right form with an argument like in Example 9.10.7.

**9.10.9. Example.** We find some geodesics on surfaces of revolution. Assume that we are given a surface of revolution in $\mathbf{R}^3$ and that the curve being revolved does not intersect the axis about which it is being revolved (something that typically would cause singularities in the surface). What we want show is that all the meridians of this surface are geodesics and a circle of latitude is a geodesics if and only if all the tangent planes at points of the circle are parallel to the axis of revolution.

**Solution.** We shall only consider the case where a regular curve $\eta(t) = (x(t),y(t))$, $t \in [a,b]$, in the plane is revolved about the x-axis and the surface has a parameterization of the form

$$\Phi(t,\theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta).$$

See Figure 9.27. We assume that $\Phi(t,\theta)$ is a regular parameterization. The proof will rely on Theorem 9.10.2 and Theorem 9.10.3(1). Now

$$\Phi_t(t,\theta) = (x'(t), y'(t)\cos\theta, y'(t)\sin\theta)$$
$$\Phi_\theta(t,\theta) = (0, -y(t)\sin\theta, y(t)\cos\theta)$$

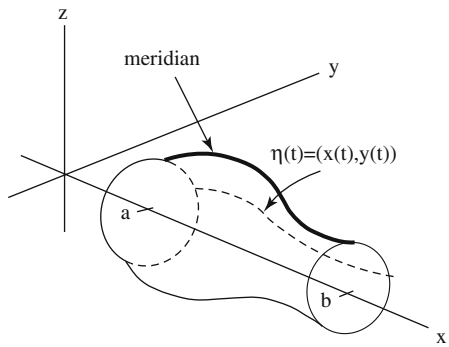



**Figure 9.27.** Geodesics on a surface of revolution.

and

$$\Phi_t(t,\theta) \times \Phi_\theta(t,\theta) = y(t)(y'(t), -x'(t)\cos\theta, -x'(t)\sin\theta).$$

The vector $\Phi_t \times \Phi_\theta$ is a surface normal. Although it may not have unit length, using it in formulas, as we shall do, rather than a normalized version will simplify expressions but not change the validity of when certain dot products are zero.

Consider the meridian

$$\alpha(t) = \Phi(t,\theta)$$

and its derivatives

$$\alpha'(t) = (x'(t), y'(t)\cos\theta, y'(t)\sin\theta)$$
$$\alpha''(t) = (x''(t), y''(t)\cos\theta, y''(t)\sin\theta)$$

It is easy to check that

$$\alpha'' \bullet ((\Phi_t \times \Phi_\theta) \times \alpha') = 0,$$

and so $\alpha(t)$ is a geodesic by the theorems mentioned.

Next, consider the circle of latitude

$$\beta(\theta) = \Phi(t,\theta)$$

and its derivatives

$$\beta'(\theta) = (0, -y(t)\sin\theta, y(t)\cos\theta)$$
$$\beta''(\theta) = (0, -y(t)\cos\theta, -y(t)\sin\theta).$$

This time

$$\beta'' \bullet ((\Phi_t \times \Phi_\theta) \times \beta') = y^3 y'.$$

But $y \neq 0$ because the curve $\eta(t)$ does not cross the x-axis, and so this expression will vanish if and only if $y' = 0$. This is equivalent to saying that $\Phi_t(t,\theta) = (x'(t),0,0)$, which is what we were trying to show.

As an application, consider the torus and its parameterization as defined in Example 9.9.18. We can think of this torus as a surface of revolution where the curve

$$\eta(t) = (0,R) + (r\cos t, r\sin t) = (r\cos t, R + r\sin t) = (x(t), y(t))$$

gets revolved about the x-axis. It is clear that the meridians, which are circles in this case, are geodesics. On the other hand, $y'(t) = \cos t = 0$ implies that $t = \pi/2$ or $3\pi/2$. In other words, the only circles of latitude that are geodesics are the inner circle with radius $R - r$ and the outer circle with radius $R + r$.

Our definition of geodesic applied only to surfaces in $\mathbf{R}^3$ because we made use of the normal vectors to the surface. Obviously, the normal curvature of a curve depends on having a normal vector, but it turns out that its geodesic curvature does not.

**9.10.10. Theorem.** (Minding) The geodesic curvature of a curve in a surface is a metric invariant.

**Proof.** Specifically, what we want to show is that the geodesic curvature depends only on the curve and the metric coefficients of the surface. How the surface is imbedded in $\mathbf{R}^3$ plays no role.

Let

$$\Phi : \mathbf{U} \to \mathbf{S}$$

be a regular one-to-one parameterization of a surface $\mathbf{S}$ in $\mathbf{R}^3$ and let $\mathbf{n}(\mathbf{p})$ be the unit normal vector to the surface $\mathbf{S}$ at the point $\mathbf{p}$. Let

$$\gamma : [a,b] \to \mathbf{S}$$

be a curve parameterized by arc-length. Express $\gamma$ in the form $\gamma(s) = \Phi(\mu(s))$, where

$$\mu : [a,b] \to \mathbf{R}^2 \quad \text{and} \quad \mu(s) = (u(s), v(s)).$$

Now

$$\gamma' = \Phi_u u' + \Phi_v v' \tag{9.69}$$

and

$$\gamma'' = \Phi_{uu} u'^2 + 2\Phi_{uv} u'v' + \Phi_{vv} v'^2 + \Phi_u u'' + \Phi_v v''. \tag{9.70}$$

Substituting the right-hand side of equations (9.58) for the $\Phi_{uu}$, $\Phi_{uv}$, and $\Phi_{vv}$ in equation (9.70) gives

$$\gamma'' = \left[Lu'^2 + 2\,Mu'v' + Nv'^2\right]\mathbf{n}$$
$$+\left[u'' + \Gamma_{11}^1 u'^2 + 2\Gamma_{12}^1 u'v' + \Gamma_{22}^1 v'^2\right]\Phi_u + \left[v'' + \Gamma_{11}^2 u'^2 + 2\,\Gamma_{12}^2 u'v' + \Gamma_{22}^2 v'^2\right]\Phi_v,$$

so that equation (9.65) clearly implies that

$$\boldsymbol{\kappa}_g n_{\mathbf{S}} = \left[u'' + \Gamma_{11}^1 u'^2 + 2\Gamma_{12}^1 u'v' + \Gamma_{22}^1 v'^2\right]\Phi_u + \left[v'' + \Gamma_{11}^2 u'^2 + 2\Gamma_{12}^2 u'v' + \Gamma_{22}^2 v'^2\right]\Phi_v.$$

To make the pattern more apparent, we change our notation slightly and rename our parameters u and v to $x_1$ and $x_2$, respectively, so that we can rewrite the last equation using summation notation as

$$\boldsymbol{\kappa}_g n_{\mathbf{S}} = \sum_{k=1}^{2}\left[x_k'' + \sum_{i,j=1}^{2}\Gamma_{ij}^k x_i' x_j'\right]\Phi_k. \tag{9.71}$$

($\Phi_1$ and $\Phi_2$ are the partials $\Phi_u$ and $\Phi_v$, respectively.) But, using Equation (9.69),

$$n_{\mathbf{S}} = \mathbf{n} \times \gamma' = \sum_{m=1}^{2} x_m'\, \mathbf{n} \times \Phi_m.$$

Therefore, if we let $c_{km} = \mathbf{n} \bullet (\Phi_k \times \Phi_m)$, then

$$\boldsymbol{\kappa}_g = (\boldsymbol{\kappa}_g n_{\mathbf{S}}) \bullet n_{\mathbf{S}} = \left(\sum_{k=1}^{2}\left[x_k'' + \sum_{i,j=1}^{2}\Gamma_{ij}^k x_i' x_j'\right]\Phi_k\right) \bullet n_{\mathbf{S}}$$
$$= \left(\sum_{k,m=1}^{2}\left[x_k'' + \sum_{i,j=1}^{2}\Gamma_{ij}^k x_i' x_j'\right] x_k' c_{km}\right).$$

Finally, since

$$c_{11} = c_{22} = 0 \quad \text{and} \quad c_{12} = -c_{21} = \sqrt{g},$$

we have shown that $\boldsymbol{\kappa}_g$ depends only on the Christoffel symbols, which are metric invariants by equation (9.59), and g. The theorem is proved.

Moving on to property (2) of straight lines, we can see from Example 9.10.8 that a geodesic in $\mathbf{S}^2$, as we have defined one, can start at a point, go round and round a great circle, and finally end up at that point again, so that it is **not** necessarily the shortest curve between any two points. On the other hand, they do satisfy property (2) locally. One can prove the following:

**9.10.11. Theorem.** Let **S** be a surface in $\mathbf{R}^3$. Let $\eta : [a,b] \to \mathbf{S}$ be a regular curve.

(1) If $\eta(t)$ is a geodesic, then there is an $\varepsilon > 0$ with the property that if $|s_1 - s_2| < \varepsilon$ then $\gamma|[s_1,s_2]$ is a curve of shortest length from $\gamma(s_1)$ to $\gamma(s_2)$.
(2) If $\eta(t)$ is a curve of shortest length from $\eta(a)$ to $\eta(b)$, then $\eta(t)$ is a geodesic.

**Proof.** For a proof of (1) using the exponential map which is defined later in this section see [Thor79]. For (2) see [MilP77].

Theorem 9.10.11 implies that the first definition of a geodesic is equivalent to the following one:

**Third definition of a geodesic:** A *geodesic* on a surface **S** in $\mathbf{R}^3$ is just a regular curve in **S** that has the property that locally it defines a curve of shortest length.

Moving on to property (3) of straight lines, it is easy to give examples that show geodesics are neither unique nor exist in general. For example, there are an infinite number of (in fact, minimal-length) geodesics between antipodal points on a sphere and the surface $\mathbf{R}^2$–**0** (a *punctured plane*) has no geodesic from (–1,0) to (1,0) (there exist curves between the two points with length arbitrarily close to 2 but none of length exactly 2). Although there may not be a geodesic between an arbitrary pair of points, geodesics **do** exist if the points are not too far apart. The proof of this fact will also show that finding geodesics is simply a matter of solving second order differential equations.

**9.10.12. Theorem.** Let **S** be a surface in $\mathbf{R}^3$. Let $\mathbf{p} \in \mathbf{S}$ and $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{S})$, $\mathbf{v} \neq 0$. Then there is an $\varepsilon > 0$ and a unique constant speed geodesic $\gamma: (-\varepsilon,\varepsilon) \to \mathbf{S}$ with $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$. (For $\mathbf{v} = 0$, the unique "geodesic" would be the constant curve $\gamma(t) = \mathbf{p}$.)

**Sketch of two proofs.** For the first proof, we use a regular parameterization $\Phi$ of a neighborhood of **p** in **S** and equation (9.71). Since we are looking for a geodesic, the left-hand side of the equation vanishes. But the partials $\Phi_1$ and $\Phi_2$ are linearly independent, being a basis of the tangent space. It follows that any solution $\gamma$ must satisfy the equations

$$x_k'' + \sum_{i,j=1}^{2} \Gamma_{ij}^k x_i' x_j' = 0, \quad k = 1,2. \tag{9.72}$$

Conversely, one can show that any such solution satisfying our initial conditions will solve our problem, namely, it must be a constant speed curve. One only has to appeal to theorems about the existence and uniqueness of solutions to differential equations to finish the proof. See [MilP77].

The proof we just sketched shows that the existence of geodesics is an intrinsic property of surfaces that does not depend on any imbedding in $\mathbf{R}^3$. The second proof does use the imbedding and normals to the surface. Its advantage is that it is more direct and useful computationally. In practice, surfaces are presented via parameterizations anyway, so computing normals is not a problem.

Near **p** one can represent the surface **S** as the zero set of some function, that is, we may assume that

$$\mathbf{S} = f^{-1}(0)$$

for some function $f: \mathbf{U} \to \mathbf{R}$, where **U** is a subset of $\mathbf{R}^3$ and the gradient of f does not vanish on **U**. In this case, the Gauss map for **S** is given by

$$\mathbf{n}(\mathbf{p}) = \frac{\nabla f}{|\nabla f|}(\mathbf{p}).$$

Now, by Theorem 9.10.3(2), $\gamma$ is a geodesic if and only if

$$\gamma''(t) = \alpha(t)\mathbf{n}(\gamma(t))$$

for some function $\alpha$. It follows that

$$\alpha = \gamma'' \bullet (\mathbf{n} \circ \gamma) = (\gamma' \bullet (\mathbf{n} \circ \gamma))' - (\gamma' \bullet (\mathbf{n} \circ \gamma)') = -\gamma' \bullet (\mathbf{n} \circ \gamma)',$$

because $\gamma' \bullet (\mathbf{n} \circ \gamma) = 0$. In other words, $\gamma$ is a geodesic if and only if

$$\gamma'' + (\gamma' \bullet (\mathbf{n} \circ \gamma)')(\mathbf{n} \circ \gamma) = 0. \tag{9.73}$$

If $\mathbf{n}(\mathbf{p}) = (n_1(\mathbf{p}), n_2(\mathbf{p}), n_3(\mathbf{p}))$ and $\gamma(t) = (x_1(t), x_2(t), x_3(t))$, then solving equation (9.73) reduces to solving the three second-order differential equations

$$\frac{d^2x_i}{dt^2} + \sum_{j,k=1}^{3} n_i(x_1,x_2,x_3)\frac{\partial n_j}{\partial x_k}(x_1,x_2,x_3)\frac{dx_j}{dt}\frac{dx_k}{dt} = 0, \quad i = 1,2,3. \tag{9.74}$$

The theorem now follows from theorems about solutions to such equations. See [Thor79] for more details. In fact it is shown there that there is a **maximal** open interval containing 0 over which the unique geodesic $\gamma(t)$ is defined.

The form of the result in Theorem 9.10.12 leads to a natural question. Can one extend the domain of the geodesic $\gamma$ in the theorem from $(-\varepsilon,\varepsilon)$ to **all** of $\mathbf{R}$?

**Definition.** A surface $\mathbf{S}$ is said to be *geodesically complete* if every geodesic $\gamma: (a,b) \to \mathbf{S}$ extends to a geodesic $\tilde{\gamma}: \mathbf{R} \to \mathbf{S}$.

**9.10.13. Example.** Neither the open unit disk nor $\mathbf{R}^2 - \mathbf{0}$ is geodesically complete.

**9.10.14. Theorem.** (The Hopf-Rinow Theorem)

(1) A surface is geodesically complete if and only if it is complete in the topological sense.
(2) If a surface is geodesically complete, then any two points can be joined by minimal-length geodesics.
(3) In a closed and compact surface there is a minimal-length geodesic between any two points.

**Proof.** See [Hick65] or [McCl97]. Part (2) implies part (3) because by Theorem 5.5.7 every compact metrizable space is complete.

Moving on to property (4) of straight lines, we first need to define what it means for a vector field along a curve to consist of parallel lines. Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and let

$$\eta : [a,b] \to \mathbf{S}$$

be a curve.

**Definition.** A *vector field along the curve* $\eta$ is a function $\mathbf{X}:[a,b] \to \mathbf{R}^3$ with the property that $\mathbf{X}(t)$ is a vector tangent to $\mathbf{S}$ at $\eta(t)$ for all t, that is, $\mathbf{X}(t) \in T_{\eta(t)}(\mathbf{S})$ for all t. The vector field $\mathbf{X}$ is said to *differentiable* if $\mathbf{X}(t)$ is a differentiable function.

Since the tangent space at a point of a surface is a vector space, it is easy to see that the set of vector fields along a curve is a vector space. Let $\mathbf{X}(t)$ be a differentiable vector field along $\eta(t)$. Although the vector $\mathbf{X}'(t)$ is not necessarily tangent to the surface at $\eta(t)$, its orthogonal projection onto the tangent space obviously will be.

**Definition.** The *covariant derivative* of $\mathbf{X}(t)$, denoted by $D\mathbf{X}/dt$, is the vector field along $\eta$ which sends $t \in [a,b]$ to the vector that is the orthogonal projection of $\mathbf{X}'(t)$ on the tangent space of $\mathbf{S}$ at $\eta(t)$. More precisely,

$$\frac{D\mathbf{X}}{dt}(t) = \mathbf{X}'(t) - (\mathbf{X}'(t) \bullet \mathbf{n}(\eta(t)))\mathbf{n}(\eta(t)),$$

where $\mathbf{n}(\eta(t))$ is a unit normal vector to $\mathbf{S}$ at $\eta(t)$.

The definition of the covariant derivative does not depend on the choice of $\mathbf{n}(\eta(t))$.

**9.10.15. Theorem.** Let $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ be differentiable vector fields along $\eta(t)$ and let $f:[a,b] \to \mathbf{R}$ be a differentiable function.

(1) $\dfrac{D}{dt}(\mathbf{X}+\mathbf{Y}) = \dfrac{D\mathbf{X}}{dt} + \dfrac{D\mathbf{Y}}{dt}$.

(2) $\dfrac{D}{dt}(f\mathbf{X}) = f'\mathbf{X} + f\dfrac{D\mathbf{X}}{dt}$.

(3) $\dfrac{d}{dt}(\mathbf{X} \bullet \mathbf{Y}) = \dfrac{D\mathbf{X}}{dt} \bullet \mathbf{Y} + \mathbf{X} \bullet \dfrac{D\mathbf{Y}}{dt}$.

**Proof.** This is a straightforward computation of derivatives (Exercise 9.10.4).

Note that if $\mathbf{S}$ is a plane, then the covariant derivative is just the ordinary derivative of the vector field. Intuitively, the covariant derivative of a vector field just measures the rate of change of the vector field as seen from "inside" the surface where one does not have any notion of a normal.

**Definition.** We say that the vector field $\mathbf{X}(t)$ on $\eta(t)$ is *parallel along* $\eta$ if $d\mathbf{X}/dt = \mathbf{0}$, that is, its covariant derivative vanishes.

One can easily show that if $\mathbf{S}$ is a plane, then the vector field $\mathbf{X}(t)$ is parallel along $\eta$ if and only if $\mathbf{X}(t)$ is constant. We see that we seem to have found a generalization of what it means for vectors to be parallel. It was Levi-Civita who introduced the covariant derivative as a means of describing parallelism. As is pointed out in

[Stok69], the origin of the definition of parallel vector fields is closely related to developable surfaces, which are defined in Section 9.15. The relevant property of these surfaces is that they can be isometrically embedded in the plane and Levi-Civita originally defined a vector field to be parallel if it mapped to a parallel vector field on the plane under such a mapping. Of course not all surfaces are developable and Levi-Civita reduced the general case to the developable case by constructing a developable surface that was tangent to the given surface along the curve $\eta$.

Here are the important properties of parallel vector fields.

**9.10.16. Theorem.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and let $\eta:[a,b] \to \mathbf{S}$ be a curve. Let $\mathbf{X}$ and $\mathbf{Y}$ be differentiable vector fields along $\eta$.

(1) If $\mathbf{X}$ is parallel along $\eta$, then the vectors of $\mathbf{X}$ have constant length.
(2) If $\mathbf{X}$ and $\mathbf{Y}$ are parallel along $\eta$, then $\mathbf{X} \bullet \mathbf{Y}$ is constant along $\eta$.
(3) If $\mathbf{X}$ and $\mathbf{Y}$ are parallel along $\eta$, then the angle between the vectors of $\mathbf{X}$ and $\mathbf{Y}$ is constant.
(4) If $\mathbf{X}$ and $\mathbf{Y}$ are parallel along $\eta$, then so are the vector fields $\mathbf{X} + \mathbf{Y}$ and $c\mathbf{X}$, for all $c \in \mathbf{R}$.

**Proof.** Part (1) follows from Theorem 9.10.15(3) and the identities

$$\frac{d}{dt}|\mathbf{X}(t)|^2 = \frac{d}{dt}(\mathbf{X}(t) \bullet \mathbf{X}(t)) = 2\frac{D\mathbf{X}}{dt} \bullet \mathbf{X}(t) = 0.$$

Part (2) also follows from Theorem 9.10.15(3), which implies that the derivative of $\mathbf{X} \bullet \mathbf{Y}$ vanishes since the covariant derivatives vanish. Part (3) follows from parts (1) and (2) and the definition of angle between vectors. Part (4) is left as an easy exercise.

**9.10.17. Theorem.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$. A constant speed regular curve $\eta:[a,b] \to \mathbf{S}$ is a geodesic according to our first definition of a geodesic if and only if the tangent vector field $\eta'(t)$ is parallel along the curve.

**Proof.** A constant speed curve $\eta$ has the property that $\eta'$ and $\eta''$ are orthogonal. In this case the covariant derivative is just the geodesic curvature.

Theorem 9.10.17 leads to the next definition of a geodesic.

**Fourth definition of a geodesic:** A *geodesic* on a surface $\mathbf{S}$ in $\mathbf{R}^3$ is just a regular curve in $\mathbf{S}$ whose tangent vectors form a parallel vector field along the curve.

Note that the new definition is equivalent to the second definition, but since it applies (by Theorem 9.10.16(1)) only to constant speed curves it is technically not equivalent to the first and third definitions, which applied to arbitrary regular curves. Overlooking this technicality, the reader may wonder why one bothers with a definition of covariant derivative and parallel vector fields when in the end the new definition of geodesic is basically the same as the one that is phrased in terms of geodesic curvature. Well, the problem is that although geodesic curvature may seem like a reasonable intuitive geometric concept, as we defined it, it is not an intrinsic invariant

of the surface but seems to depend on the surface being imbedded in $\mathbf{R}^3$ because we used a normal vector to the surface. We shall see later in Section 9.16 that geodesics really do not depend on any imbedding. In fact, they can be defined for any manifold, not just a surface, and depend only on the metric coefficients. It is at that point of abstraction that one sees why the preferred development of geodesics is via a global notion of parallel vector fields and why the geodesic curvature is the secondary notion. Of course, one will have to give a more intrinsic definition of parallel vector fields since the definitions for the covariant derivative and parallel vector fields in this section also depended on a normal vector.

As one final point about parallel vector fields, we address the question of their existence and uniqueness.

**9.10.18. Theorem.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and let $\eta:[a,b] \to \mathbf{S}$ be a curve. Let $\mathbf{v} \in T_{\eta(a)}(\mathbf{S})$. Then there exists a unique parallel vector field $\mathbf{X}(t)$ along $\eta$ with $\mathbf{X}(a) = \mathbf{v}$.

**Proof.** The proof simply consists of analyzing the condition for a vector field to be parallel. The condition is equivalent to finding a solution to first order differential equations with initial condition $\mathbf{v}$. See [Thor79].

An interesting consequence of the last theorem is

**9.10.19. Corollary.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and let $\gamma(s)$ be a geodesic in $\mathbf{S}$. Then a vector field $\mathbf{X}$ along $\gamma$ is parallel along $\gamma$ if and only if both the length of all the vectors of $\mathbf{X}$ and the angle between them and $\gamma'(s)$ is constant.

**Proof.** See [Thor79].

Furthermore, since parallel vector fields are completely defined by their initial vector, one introduces the following useful terminology.

**Definition.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$ and let $\eta:[a,b] \to \mathbf{S}$ be a curve. Let $\mathbf{v} \in T_{\eta(a)}(\mathbf{S})$. Let $\mathbf{X}$ be the unique vector field along $\eta$ that satisfies $\mathbf{X}(a) = \mathbf{v}$. If $c \in [a,b]$, then $\mathbf{X}(c)$ is called the *parallel translate* of $\mathbf{v}$ to $\eta(c)$ along $\eta$.

We finish our introduction to geodesics on surfaces by defining one more important map. It turns out that one can rephrase the result in Theorem 9.10.12 as follows:

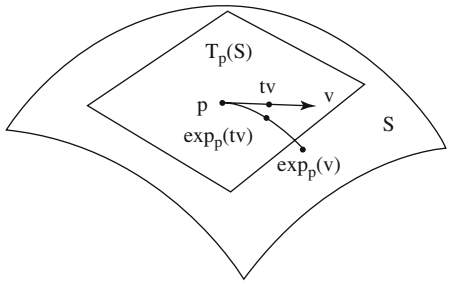
**9.10.20. Theorem.** Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$. If $\mathbf{p} \in \mathbf{S}$, then there is an $\varepsilon_{\mathbf{p}} > 0$ so that for all $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{S})$ with $|\mathbf{v}| < \varepsilon_{\mathbf{p}}$ there is a unique geodesic $\gamma:(-2,2) \to \mathbf{S}$ with $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$.

**Proof.** Exercise 9.10.5.

Using the notation in Theorem 9.10.20 let

$$\mathbf{U}(\mathbf{p}) = \{\mathbf{v} \in T_{\mathbf{p}}(\mathbf{S}) \mid |\mathbf{v}| < \varepsilon_{\mathbf{p}}\} \subseteq T_{\mathbf{p}}(\mathbf{S})$$

**Figure 9.28.** The exponential map.




and define a map

$$\exp_{\mathbf{p}} : \mathbf{U}(\mathbf{p}) \to \mathbf{S} \quad \text{by} \quad \exp_{\mathbf{p}}(\mathbf{v}) = \gamma(1).$$

See Figure 9.28.

**Definition.** The map $\exp_{\mathbf{p}}$ is called the *exponential map* of the surface $\mathbf{S}$ at $\mathbf{p}$.

The exponential map is defined in a neighborhood of the origin of each tangent space. If the surface is geodesically complete, then the exponential map is defined on all of $T_{\mathbf{p}}(\mathbf{S})$. The exponential maps $\exp_{\mathbf{p}}$ at points $\mathbf{p}$ define a map

$$\exp: \begin{pmatrix} \text{open neighborhood of zero cross-section} \\ \text{in total space of tangent bundle } \tau_{\mathbf{S}} \end{pmatrix} \to \mathbf{S}$$

that is also called the *exponential map*.

**9.10.21. Theorem.**

(1) Both $\exp_{\mathbf{p}}$ and exp are differentiable maps.
(2) The map $\exp_{\mathbf{p}}$ is a diffeomorphism of a neighborhood of the origin in $T_{\mathbf{p}}(\mathbf{S})$ onto a neighborhood of $\mathbf{p}$ in $\mathbf{S}$.
(3) For each $\mathbf{p} \in \mathbf{S}$ and $\mathbf{v} \in \mathbf{U}(\mathbf{p})$, the unique geodesic $\gamma$ from $\mathbf{p}$ to $\mathbf{q} = \exp_{\mathbf{p}}(\mathbf{v})$ is defined by

$$\gamma(t) = \exp_{\mathbf{p}}(t\mathbf{v}).$$

**Proof.** See [Thor79].

We see from Theorem 9.10.21 that the exponential map essentially formalizes the concept of an "orthogonal projection" of a neighborhood of the origin in the tangent space $T_{\mathbf{p}}(\mathbf{S})$ to $\mathbf{S}$.

## 9.11 Envelopes of Surfaces

Although the concept of envelopes of surfaces was often used by geometers it is actually tricky to define carefully. We shall simply extend the definition given in the case of curves.

**Definition.** Let $\alpha_t : [0,1] \to \mathbf{R}^3$ be a one-parameter family of surfaces defined by

$$\alpha_t(u,v) = \alpha(u,v,t)$$

for some $C^\infty$ function

$$\alpha : [0,1] \times [0,1] \times [0,1] \to \mathbf{R}^3.$$

An *envelope* of this family is defined to be a surface p(u,v) that is not a member of this family but that is tangent to some member of the family at every point.

Spivak ([Spiv75]) discusses the envelope of a family of planes. He shows that "in general" a one-parameter family of planes has an envelope that is either a generalized cylinder, a generalized cone, or the tangent surface of a curve. He also explains how this led to a definition of parallel translation along a Riemannian manifold.

## 9.12 Canal Surfaces

This section is on a special type of envelope surface that is relevant to CAGD.

**Definition.** A *canal surface* is the envelope of a one-parameter family of spheres S(t). If $\gamma(t)$ is the center of the sphere S(t), then the curve $\gamma(t)$ is called the *center curve* of the canal surface. The function r(t), where r(t) is the radius of the sphere S(t), is called the *radius function* of the canal surface. A canal surface whose radius function is constant is called a *tube surface*.

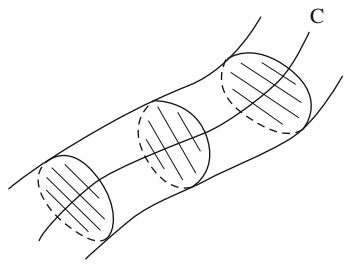
Canal surfaces were first defined and studied by Gaspard Monge. If the center curve for a canal surface is a straight line, then we get a surface of revolution. In general, canal surfaces are a type of "sweep" surface. They are the boundary of the solid that one gets by sweeping a sphere along a curve.

**9.12.1. Lemma.** If S(t) is the one-parameter family of spheres that defines a canal surface **S**, then $S(t) \cap \mathbf{S}$ is a circle for every t.

**Proof.** This follows from the fact that $S(t) \cap \mathbf{S}$ is the limit of the intersections $S(t - \delta) \cap S(t + \delta)$ as $\delta$ approaches 0.

**Definition.** The circles $S(t) \cap \mathbf{S}$ in Lemma 9.12.1 are called the *characteristic circles* of the canal surface.

**Figure 9.29.** A tube surface.




It follows from Lemma 9.12.1 that canal surfaces can also be thought of as a union of circles with centers on a curve. Note that since both S(t) and **S** are tangent on $S(t) \cap \mathbf{S}$, they have the same normals along their intersection. Therefore in the special case of a tube surface, a characteristic circle is the boundary of a disk that intersects the surface orthogonally. This leads to yet another view of tube surfaces, namely, as the boundary of a solid that is obtained by sweeping a disk of constant radius orthogonally along a curve **C**. See Figure 9.29.

Before we state the next theorem, recall from Section 9.9 that the centers of curvatures at a point **p** of a surface **S** lie along the normal line to the surface at that point. These centers lie on segments in the normal line whose endpoints were called the focal points of the surface **S**, and the set of these was called the surface of centers (or the evolute).

**9.12.2. Theorem.**

(1) The characteristic circles of a canal surface are lines of curvature for the surface.
(2) The center curve for a canal surface consists of centers of curvature for the surface, so that one component of its surface of centers is a curve.
(3) Every surface with the property that one component of its surface of centers is a curve is a canal surface.

**Proof.** See [Gray98].

**9.12.3. Theorem.** Let **S** be a canal surface with radius function r(t) and center curve $\gamma:[a,b] \to \mathbf{R}^3$ whose tangent vectors $\gamma'(t)$ have unit length and whose curvature is nonzero. Then **S** admits a parameterization of the form

$$\phi(t,\theta) = \gamma(t) + r(t)\left[-r'(t)T(t) \pm \sqrt{1-r'(t)^2}\,(-\cos\theta\, N(t) + \sin\theta\, B(t))\right],$$

where (T(t),N(t),B(t)) is the Frenet frame of the curve $\gamma(t)$.

**Proof.** See [Gray98].

If the radius of the spheres or disks for a tube surface is small enough we get an immersed surface. We can relate the curvature of the center curve and the Gauss curvature of the surface.

**9.12.4. Theorem.** Let $\boldsymbol{\kappa}(s)$ be the curvature function of a space curve $\gamma:[a,b] \to \mathbf{R}^3$. If K is the Gauss curvature of a canal surface **S** that has constant radius function and center curve $\gamma$, then

$$\int_a^b \boldsymbol{\kappa}(s)ds = \frac{1}{4}\int_{\mathbf{S}} |K|\, dA.$$

**Proof.** See [Spiv75].

## 9.13 Involutes and Evolutes of Surfaces

The discussion of involutes and evolutes of curves extends to surfaces in interesting ways. We already defined the evolute (or surface of centers) of a surface in Section 9.9. It also played a role in Theorem 9.12.3. The evolute consists of the set of foci on the normal lines at the points of the surface. The foci will be the same point if and only if **p** is an umbilic point. Furthermore, Theorem 9.10.10 implies that one of those points will be at infinity if and only if the Gauss curvature vanishes at **p**. The common name, "surface of centers," for the evolute suggests that it is a surface, but that is not necessarily the case. For a sphere, the evolute is just the center of the sphere, but one can show that the sphere is the only surface for which the evolute is a point. Theorem 9.12.3 shows that some surfaces have evolutes that consist of curves, but one can describe all such surfaces. The only surfaces with this property were discovered by the French mathematician Dupin in 1822 ([Dupi22]).

**Definition.** A *Dupin cyclide* or simply *cyclide* is a surface whose evolute consists of two curves. The two curves are called the *spines* of the cyclide.

Figure 9.30 shows a "ring" cyclide which is a kind of torus, but rather than rotating a circle of fixed radius about another circle the radius of the rotating circle varies. A cyclide is a special case of a canal surface where only **one** sheet of the evolute was required to be a curve. Cyclides are very interesting surfaces and can be defined in a number of different ways ranging from very geometric constructions to explicit formulas. We shall discuss a few of their major geometric properties. Good references for more details are [ChDH89], and [Prat90], [Boeh90], and [HilC99].

First of all, the definition of a cyclide we have given is due to Maxwell ([Maxw68]) and is not Dupin's original definition.

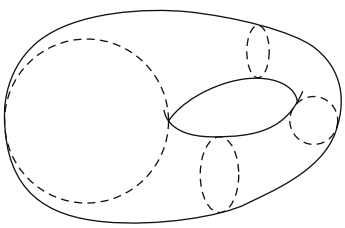


**Figure 9.30.** A ring cyclide.


**Figure 9.31.** A cyclide as an envelope of spheres.

**Dupin's definition** ([Dupi22])**:** A *cyclide* is the envelope of spheres that touch three fixed spheres.

The example in Figure 9.31 captures Dupin's basic idea. Figure 9.31(a) shows four spheres $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, and $\mathbf{S}$ with centers in the x-y plane. The sphere $\mathbf{S}$ touches the other three. Imagine raising the center of the sphere $\mathbf{S}$ in the z-direction while at the same time letting its radius expand so that we maintain contact with the other three spheres. We keep raising the center in this way until the radius becomes "infinity." At that point, we wrap around to "minus infinity" and let the radius decrease so that the z-coordinate of the center now moves from minus infinity to zero. Note that in the first stage the spheres $\mathbf{S}_i$ were outside of $\mathbf{S}$ but now they are inside. See Figure 9.31(b). The envelope of the moving sphere $\mathbf{S}$ is our cyclide. See [ChDH89] for a more thorough discussion of this process and why Maxwell's and Dupin's definition of a cyclide are equivalent.

**Definition.** A pair of conics is said to be an *anticonic pair* or *anticonics* if they lie in orthogonal planes and the vertices of one are the foci of the other and vice versa.

**9.13.1. Theorem.** The two spines of a cyclide are conics. In fact, they are an anticonic pair.

**Proof.** See [ChDH89].

There is a nice geometric construction using a string on which Maxwell's definition is based. Figure 9.32(a) shows the construction in the special case of a torus. We tie a string of fixed length to the center $\mathbf{A}$ of a planar circle and then wrap the string around the circle. The torus will consist of all points swept out by the endpoint $\mathbf{B}$ of the string assuming that we keep it taut and there are no lateral forces at the point $\mathbf{P}$ where the string touches the circle. It is easy to see that the normal lines to the surface through $\mathbf{P}$ and $\mathbf{B}$ will either intersect the vertical axis $\mathbf{L}$ or be parallel to it. Figure 9.32(b) shows the construction for a general cyclide. Consider an ellipse $\mathbf{E}$ in the x-y plane with foci at $(\pm c,0,0)$ and vertices at $(\pm a,0,0)$ and $(0,\pm b,0)$. Let $\mathbf{H}_1$ and $\mathbf{H}_2$ be branches of the hyperbola $\mathbf{H}$ in the x-z plane whose foci and vertices are the vertices and foci of the ellipse $\mathbf{E}$, respectively. Tie a string to the focus $\mathbf{A} = (-c,0,0)$ of $\mathbf{E}$ and

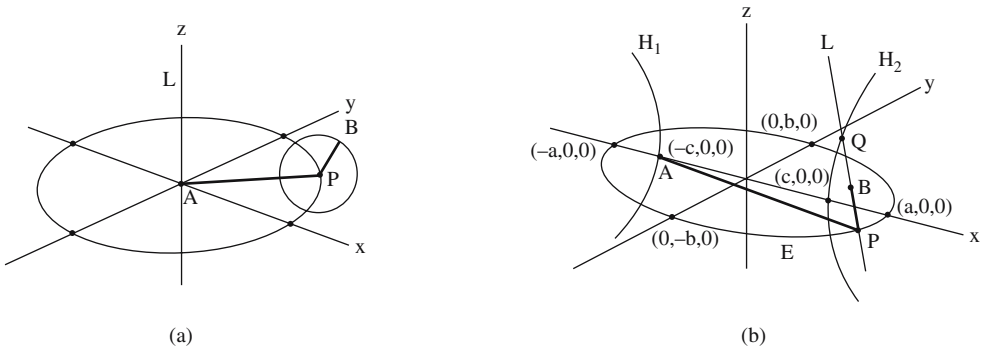



**Figure 9.32.** A string construction for a cyclide.

wrap it around **E**. The cyclide is the locus of points swept out by the end point **B** of the string as we keep it taut and ensure that there are no lateral forces at the point **P** where it touches **E**. The line **L** through **P** and **B** will intersects **H**.

**9.13.2. Theorem.** Cyclides are the only surfaces all of whose lines of curvature are circles.

**Proof.** See [HilC99].

It is useful to divide the cyclides into different types. Recall that the spines of cyclides are anticonics.

**Definition.** A *central*, *parabolic*, or *revolute cyclide* is a cyclide whose spines are an ellipse/hyperbola, parabola/parabola, or circle/straight line pair, respectively. A *degenerate cyclide* is a cyclide at least one of whose spines is a degenerate conic.

For a finer subdivision of each cyclide type into "horned," "ring," and "spindle" cyclides see [ChDH89].

Now, in general, the evolute of a surface **S** consists of two sheets of surface. One can show that the normals to **S** are tangent to the evolute at the foci. Therefore the normals to **S** are the common tangents to the two sheets of the evolute. One can turn this around.

**Definition.** Given two arbitrary surfaces $\mathbf{S}_1$ and $\mathbf{S}_2$, a surface **S** whose normals define a family of lines that are tangent to both of the given surfaces is called the *involute* of the surfaces.

The involute **S** has the property that the two surfaces $\mathbf{S}_1$ and $\mathbf{S}_2$ are the evolute of it. One can show that a necessary and sufficient condition for this to happen is that the tangent planes at the points where a normal of **S** touches $\mathbf{S}_1$ and $\mathbf{S}_2$ must be orthogonal. A pair of confocal quadrics of unlike type is an example of this. See [HilC99].

## 9.14 Parallel Surfaces

The main reason for this section on parallel surfaces is that, like parallel curves, the topic is important for CAGD.

Let p(u,v) be a regular parameterization for a surface **S** in $\mathbf{R}^3$. Let

$$N(u,v) = \frac{\partial p}{\partial u}(u,v) \times \frac{\partial p}{\partial v}(u,v). \tag{9.75}$$

Since we have a regular parameterization, N(u,v) is nonzero and we can normalize it to be of unit length. If

$$\mathbf{n}(u,v) = \frac{N(u,v)}{|N(u,v)|}, \tag{9.76}$$

then $\mathbf{n}(u,v)$ is a unit normal vector to **S** at p(u,v).

**Definition.** A *parallel surface* to p(u,v) is a surface $\mathbf{S}_d$ with a parameterization $p_d(u,v)$ of the form

$$p_d(u,v) = p(u,v) + d\mathbf{n}(u,v) \tag{9.77}$$

where d is any nonzero real number. In CAGD a parallel surface is called an *offset surface*.

It follows from equation (9.77) that

$$(p_d)_u = p_u + d\mathbf{n}_u \quad \text{and} \quad (p_d)_v = p_v + d\mathbf{n}_v. \tag{9.78}$$

**Assumption:** In the rest of this section we shall assume that $p_u$ and $p_v$ are principal directions!

There is no loss of generality with this assumption because the quantities we shall want to compute are independent of parameterizations and by Theorem 9.9.22 such parameterizations always exist. The assumption will greatly simplify our computations because F = M = 0 in this case by Theorem 9.9.23.

Substituting $\mathbf{n}_u$ and $\mathbf{n}_v$ from the Weingarten equations with F = M = 0 into equations (9.78) and using Theorem 9.9.25 gives

$$(p_d)_u = p_u\left(1 - d\frac{L}{E}\right) = p_u(1 - d\kappa_1) \tag{9.79a}$$

and

$$(p_d)_v = p_v\left(1 - d\frac{N}{G}\right) = p_v(1 - d\kappa_2), \tag{9.79b}$$

where $\kappa_i$ are the principal normal curvatures of $\mathbf{S}$. Let $E_d$, $F_d$, and $G_d$ be the coefficients of the first fundamental form for $\mathbf{S}_d$. Their definition and equations (9.79) imply that

$$E_d = E(1 - d\kappa_1)^2, \quad F_d = 0, \quad G_d = G(1 - d\kappa_2)^2. \tag{9.80}$$

Next, let $N_d(u,v)$ be the normal vector to the surface $\mathbf{S}_d$ at $p_d(u,v)$ and

$$\mathbf{n}_d(u,v) = \frac{N_d(u,v)}{|N_d(u,v)|}. \tag{9.81}$$

Then again using equations (9.79),

$$N_d(u,v) = (p_d)_u(u,v) \times (p_d)_v(u,v) = \left(1 - (\kappa_1 + \kappa_2)d + \kappa_1\kappa_2 d^2\right)N(u,v)$$

that is,

**9.14.1. Theorem.**

$$N_d(u,v) = \left(1 - 2\,Hd + Kd^2\right)N(u,v), \tag{9.82}$$

where K and H are the Gauss and mean curvatures of the surface $\mathbf{S}$ at p(u,v), respectively.

Since the expression $1 - 2Hd + Kd^2$ in equation (9.82) factors into $(1 - d\kappa_1)(1 - d\kappa_2)$, we always need to choose a d so that neither of these factors is zero. In fact, the map

$$\begin{aligned} \mathbf{S} &\to \mathbf{S}_d \\ p(u,v) &\to p(u,v) + d\mathbf{n}(u,v) \end{aligned}$$

is then one-to-one in a neighborhood of p(u,v). Furthermore, note that the parallel surface could be oriented in the opposite direction from the original surface. That happens when

$$1 - 2Hd + Kd^2 < 0.$$

Define $\sigma$ to be the sign of that expression, that is,

$$\mathbf{n}_d = \sigma\,\mathbf{n}. \tag{9.83}$$

Next, the definition of the coefficients $L_d$, $M_d$, and $N_d$ of the second fundamental form of $\mathbf{S}_d$ and the formulas above imply that

$$L_d = \sigma(1 - d\kappa_1)L, \quad M_d = 0, \quad N_d = \sigma(1 - d\kappa_2)N. \tag{9.84}$$

It follows from Theorem 9.9.23 that $F_d = M_d = 0$ and $(p_d)_u$ and $(p_d)_v$ are the principal directions for $S_d$. These values and Theorems 9.9.15, 9.9.16, and 9.9.25 prove

**9.14.2. Theorem.**

(1) The principal normal curvatures of $\mathbf{S}_d$ are

$$\frac{\sigma\kappa_1}{1 - d\kappa_1} \quad \text{and} \quad \frac{\sigma\kappa_2}{1 - d\kappa_2}.$$

(2) The Gauss and mean curvatures of $\mathbf{S}_d$ are

$$K_d = \frac{K}{1 - 2Hd + Kd^2} \quad \text{and} \quad H_d = \sigma\frac{H - Kd}{1 - 2Hd + Kd^2},$$

respectively.

## 9.15 Ruled Surfaces

An interesting special class of surfaces is the class of ruled surfaces. These are the surfaces obtained by sweeping a straight line along a curve.

**Definition.** A *ruled surface* is a surface **S** that admits a parameterization of the form

$$p(u,v) = q(u) + v\,\alpha(u), \tag{9.85}$$

where $q(u)$ and $\alpha(u)$ are curves called the *base curve* and *directrix* of **S**, respectively, with $|\alpha(u)| = 1$. If u is fixed, then $p(u,v)$ generates a line $\mathbf{L}_u$ that passes through $q(u)$ and has unit direction vector $\alpha(u)$. These lines are called the *rulings* of **S**. If $\alpha(u)$ is constant, then **S** is called a *cylinder*. If $\alpha'(u)$ never vanishes, then **S** is called a *non-cylindrical surface*. A ruled surface that admits two distinct rulings is called a *doubly ruled surface*.

**9.15.1. Example.** Ordinary cylinders are ruled surfaces. For example, the cylinder of radius 1 centered on the z-axis can be parameterized by

$$p(u,v) = (\cos u, \sin u, 0) + v(0,0,1).$$

More generally, if we sweep a vertical line along a curve **C** in the plane, we get what are called *generalized cylinders*. See Figure 9.33(a). Cones are also ruled surfaces because they are obtained by sweeping lines that all intersect in a fixed point **p** along some curve. See Figure 9.33(b).

**9.15.2. Example.** Consider the surface obtained by sweeping rotating lines along the z-axis that remain parallel to the xy-plane and that is parameterized by

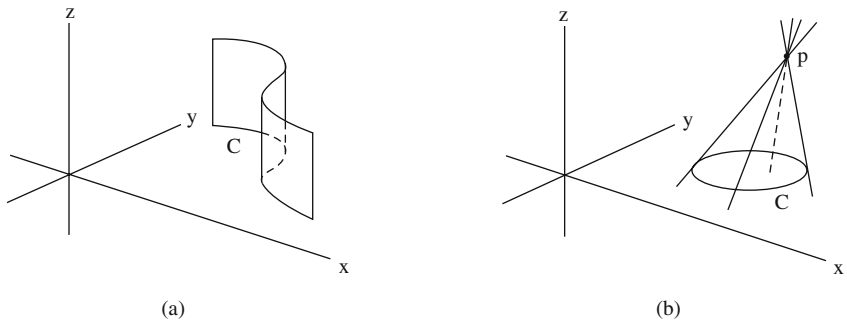



**Figure 9.33.** Ruled surfaces.

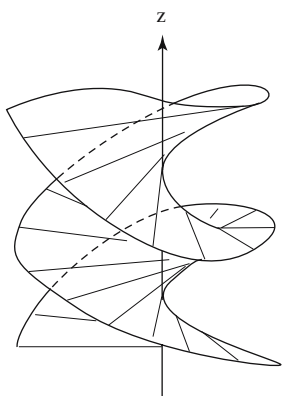



**Figure 9.34.** Sweeping a rotating line along the z-axis.

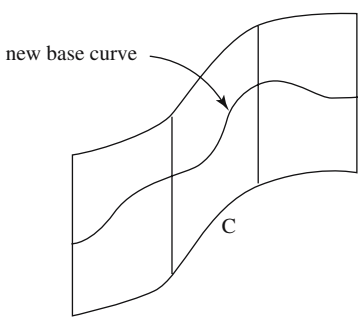



**Figure 9.35.** A ruled surface with general base curve.

$$p(u,v) = (v\cos u, v\sin u, bu), \quad b \neq 0.$$

This surface is called a *helicoid*. See Figure 9.34.

Note that, in a sense, there is nothing special about the base curve q(u) of a ruled surface. We could choose any curve that meets each line in the ruling in precisely one point as base curve and the resulting ruled surface would be the same. See Figure 9.35.

Our next goal is to compute the Gauss curvature of a ruled surface. By Theorem 9.9.15 we need merely compute the functions E, F, G, L, M, and N. Now

$$p_u = q' + v\alpha', \quad p_v = \alpha, \tag{9.86a}$$

$$p_{uu} = q'' + v\alpha'', \quad p_{uv} = \alpha', \quad p_{vv} = 0. \tag{9.86b}$$

If

$$\mathbf{n}(u,v) = \frac{p_u \times p_v}{|p_u \times p_v|} \tag{9.87}$$

is the unit normal to **S** at p(u,v), then

$$\begin{aligned} L &= \mathbf{n} \bullet p_{uu} \\ M &= \mathbf{n} \bullet p_{uv} = \mathbf{n} \bullet \alpha' \\ N &= \mathbf{n} \bullet p_{vv} = 0. \end{aligned}$$

Since N is zero, $LN - M^2 = -M^2$, and so we do not need to compute L. To make the rest of the computations easier, we replace q(u) with another base curve.

If $\alpha'(u) \neq 0$, then for small $\varepsilon$ the ruling $\mathbf{L}_{u+\varepsilon}$ will not be parallel to the ruling $\mathbf{L}_u$ and so there will be a unique point $\mu(u,\varepsilon)$ on $\mathbf{L}_u$ that is closest to $\mathbf{L}_{u+\varepsilon}$. Let

$$\mu(u) = \lim_{\varepsilon \to 0} \mu(u,\varepsilon).$$

**Definition.** The curve $\mu(u)$ is called the *striction curve* or *line of striction* for the ruled surface **S**. Its points are called the *central points* of the ruled surface.

The striction curve clearly does not depend on the parameterization of the surface.

**9.15.3. Lemma.** The curve $\mu(u)$ is well defined and satisfies $\mu' \bullet \alpha' = 0$.

**Proof.** Exercise 9.15.1.

From now on we assume that our original base curve was the striction curve for the surface.

Next, notice that $\alpha \bullet \alpha = 1$ implies

$$\alpha \bullet \alpha' = 0. \tag{9.88}$$

Equation (9.88) and Lemma 9.15.3 imply that $q' \times \alpha$ is parallel to $\alpha'$, that is,

$$q' \times \alpha = \lambda \alpha',$$

for some function $\lambda$. Dotting both sides with $\alpha'$ gives that

$$\lambda = \frac{(q' \times \alpha) \bullet \alpha'}{\alpha' \bullet \alpha'}. \tag{9.89}$$

**Definition.** The function $\lambda(u)$ is called the *distribution parameter* of the ruled surface.

Note that the distribution parameter is constant along each ruling.
If we use Lemma 9.15.3 again and also equation (9.88) we have that

$$\begin{aligned} E &= p_u \bullet p_u = (q' + v\alpha') \bullet (q' + v\alpha') = q' \bullet q' + v^2\, \alpha' \bullet \alpha' \\ F &= p_u \bullet p_v = (q' + v\alpha') \bullet \alpha = q' \bullet \alpha \\ G &= p_v \bullet p_v = 1. \end{aligned}$$

From this, the vector identity in Proposition 1.10.4(4), and the definition of $\lambda$ we get

$$\begin{aligned} EG - F^2 &= v^2\alpha' \bullet \alpha' + q' \bullet q' - (q' \bullet \alpha)^2 \\ &= v^2\alpha' \bullet \alpha' + (q' \bullet q')(\alpha \bullet \alpha) - (q' \bullet \alpha)^2 \\ &= v^2\alpha' \bullet \alpha' + (q' \times \alpha) \bullet (q' \times \alpha) \\ &= \left(\lambda^2 + v^2\right)(\alpha' \bullet \alpha') \end{aligned}$$

Furthermore, equations (9.86), (9.87), and the fact that $(\alpha' \times \alpha) \bullet \alpha' = 0$ implies

$$M^2 = (\mathbf{n} \bullet \alpha')^2 = \frac{(((q' + v\alpha') \times \alpha) \bullet \alpha')^2}{|p_u \times p_v|^2} = \frac{((q' \times \alpha) \bullet \alpha')^2}{\left(\lambda^2 + v^2\right)(\alpha' \bullet \alpha')} = \frac{\lambda^2(\alpha' \bullet \alpha')}{\lambda^2 + v^2}.$$

We collect all these facts together in the theorem we were after.

**9.15.4. Theorem.** The Gauss curvature of the ruled surface defined by equation (9.85) with $\alpha' \neq 0$ is given by

$$K = K(u,v) = -\frac{M^2}{EG - F^2} = -\frac{\lambda^2}{\left(\lambda^2 + v^2\right)^2},$$

where $\lambda$ is defined by equation (9.89).

It follows from the theorem that the Gauss curvature of a ruled surface is $\leq 0$.

**Definition.** A ruled surface is called a *developable surface* if the tangent plane is constant along each ruling.

Developable surfaces have also been defined as the envelope of a one-parameter family of planes. What this means is that the points $p(u,v)$ of the surface satisfy an equation of the form

$$p(u,v) \bullet N(t) = q(t), \tag{9.90}$$

for a nonzero vector-values function $N(t)$ and a real function $q(t)$.

**9.15.5. Theorem.** If **S** is a surface in $\mathbf{R}^3$, then the following statements about **S** are equivalent:

(1) **S** is developable.
(2) **S** is an envelope of a one-parameter family of planes, that is, the points of **S** satisfy an equation of the form (9.90).
(3) **S** is a ruled surface described by an equation of the form (9.85) for which $(q' \times \alpha) \bullet \alpha' = 0$. (If $\alpha' \neq 0$, then the last condition is equivalent to saying that the distribution parameter vanishes.)
(4) The Gauss curvature of **S** is zero.
(5) **S** can be isometrically imbedded in the plane.

**Proof.** See [Laug65], [Lips69], [Spiv75], or [Stok69]. Note that Theorem 9.10.24 and (4) implies (5).

Theorem 9.15.5 explains the terminology "developable" because historically these surfaces are the ones one said could be "developed" or "rolled out" on (that is, isometrically imbedded in) the plane.

**9.15.6. Theorem.** A necessary and sufficient condition that a curve on a surface is a line of curvature is that the surface normals along this curve form a developable surface.

**Proof.** This is a theorem of Monge. See [Will59].

## 9.16 The Cartan Approach: Moving Frames

Our development of differential geometry has so far basically followed the classical approach based on local coordinates and parameterizations. In this section we would like to sketch what a modern development would look like and discuss Cartan's theory of moving frames.

Looking back at curves in $\mathbf{R}^3$, it turns out that a key element behind the Serret-Frenet formulas, which is more fundamental than the actual formulas that one gets, is the realization that relating the derivative of the Frenet frame (which measures how it changes) to the Frenet frame uncovers a lot of geometric invariants. In the case of surfaces, its geometry was captured by the way that the basis $(\Phi_u, \Phi_v, \mathbf{n})$ changed for a parameterization $\Phi$, where $\mathbf{n}$ was the normal vector $\Phi_u \times \Phi_v$ normalized to unit length. This suggests it might be useful to look at frame fields (defined below) and how they change in general. To keep the discussion as simple as possible we shall restrict our definitions to orthonormal bases right from the start. There is no loss in generality because the Gram-Schmidt algorithm would let us replace arbitrary bases with orthonormal ones anyway. Furthermore, one would move to those eventually in any case because one needs the normalization inherent in orthonormal bases if one wants to get geometric invariants. (Every basis of a vector space is an orthonormal basis with respect to some inner product on that vector space and we are interested in geometric properties with respect to a given metric.) We start with definitions for Euclidean space.

We shall be talking a lot about vector fields in this section. Although we mostly deal with curves and surfaces in $\mathbf{R}^3$, we again want to do things in such a way so that

generalizations to n-dimensional submanifolds $\mathbf{M}^n$ of $\mathbf{R}^k$ are clear. Up to now we have used the definition of tangent vectors and tangent planes as presented in Section 8.4. In this section it will be convenient to use a variant of those definitions, one that is implicit in equation (8.30) in Section 8.10. Tangent vectors will now be associated to points. The formal way to do this was to define the tangent space $T_{\mathbf{p}}(\mathbf{M}^n)$ at $\mathbf{p} \in \mathbf{M}^n$ by

$$T_{\mathbf{p}}(\mathbf{M}^n) = \{(\mathbf{p},\mathbf{v}) \mid \mathbf{v} \text{ is a tangent vector (in the sense of Section 8.4) to } \mathbf{M}^n \text{ at } \mathbf{p}\} \subseteq \mathbf{M} \times \mathbf{R}^k.$$

There is one caveat, however. If we were to always use this pair $(\mathbf{p},\mathbf{v})$ notation for a tangent vector, even if we were to abbreviate it somewhat to $\mathbf{v}_{\mathbf{p}}$ as we shall do at times, the readability of many of expressions and equations in the rest of this section would suffer greatly unfortunately. Things would look much more complicated than they really are. Therefore, we shall be sloppy and use "$\mathbf{v}$" to denote either the pair $(\mathbf{p},\mathbf{v})$ or $\mathbf{v}$ itself whenever it is convenient. Sometimes an expression may contain two instances of "$\mathbf{v}$" where each has a different meaning. The reader will know which is being meant at any given time because there will only be one obvious correct meaning. To help out making the distinction, we shall, when possible, refer to the "tangent vector $\mathbf{v}$" or the "tangent vector $\mathbf{v}$ at $\mathbf{p}$" when we mean $(\mathbf{p},\mathbf{v})$ and simply the "vector $\mathbf{v}$" if we do mean just $\mathbf{v}$. By the way, the only reason that we run into this notational problem is so that the reader can see the close connection between abstract concepts, such as tangent vectors, with "ordinary" functions and derivatives for Euclidean space.

Specializing to the case $\mathbf{M} = \mathbf{R}^k$, let $\mathbf{X}$ be a vector field of $\mathbf{R}^k$ defined over some open subset $\mathbf{A}$. We can write $\mathbf{X}$ in the form

$$\mathbf{X}(\mathbf{p}) = (\mathbf{p}, g(\mathbf{p})),$$

for some function $g: \mathbf{A} \to \mathbf{R}^k$. Given a curve $\gamma: [a,b] \to \mathbf{A}$, define

$$\mathbf{X}'_\gamma(\gamma(t)) = \left(\gamma(t), \frac{d}{dt} g(\gamma(t))\right) \tag{9.91}$$

This definition of course only makes sense when $\gamma$ is one-to-one but everything in this section involves local properties and so assuming that parametric curves are locally one-to-one is not a problem.

**Definition.** The vector field $\mathbf{X}'_\gamma$ is called the *derivative of* $\boldsymbol{X}$ *along the curve* $\gamma$.

Now let $\mathbf{X}$ be a vector field of $\mathbf{R}^3$ defined in a neighborhood of $\mathbf{p}$ and let $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{R}^3)$.

**Definition.** The *covariant derivative*, of $\mathbf{X}$ at $\mathbf{p}$ with respect to $\mathbf{v}$, denoted by $\nabla_{\mathbf{v}}\, \mathbf{X}(\mathbf{p})$, is the tangent vector of $\mathbf{R}^3$ at $\mathbf{p}$ defined by

$$\nabla_{\mathbf{v}} \mathbf{X}(\mathbf{p}) = \mathbf{X}'_\gamma(\mathbf{p}),$$

where $\gamma$ is the curve $\gamma(t) = \mathbf{p} + t\mathbf{v}$.

The covariant derivative is a kind of directional derivative and the reader may wonder why we do not call it just that. The reason is that this is a special case of a concept that is defined for abstract manifolds. It is the classical terminology that was motivated by the same reason that covariant tensors are called what they are.

The next lemma shows that the covariant derivative can be computed from any curve $\gamma$ through $\mathbf{p}$ having tangent vector $\mathbf{v}$ at $\mathbf{p}$, not just the special curve in the definition.

**9.16.1. Lemma.** Let $\mathbf{p} \in \mathbf{R}^3$ and $\gamma:[a,b] \to \mathbf{R}^3$ be any curve with $\gamma(c) = \mathbf{p}$ and $\gamma'(c) = \mathbf{v}$. If $\mathbf{X}$ is a vector field of $\mathbf{R}^3$ defined in a neighborhood of $\mathbf{p}$, then

$$\nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p}) = \mathbf{X}'_\gamma(\mathbf{p}).$$

**Proof.** Using the notation of Equation (9.91), if

$$g(\mathbf{q}) = (g_1(\mathbf{q}), g_2(\mathbf{q}), \ldots, g_k(\mathbf{q})),$$

then

$$\begin{aligned}\mathbf{X}'_\gamma(\mathbf{p}) &= \left(\mathbf{p}, \left(\frac{d}{dt}g(\gamma(t))\right)(c)\right)\\ &= (\mathbf{p}, Dg(\mathbf{p})(D\gamma(c)))\\ &= (\mathbf{p}, (\nabla g_1(\mathbf{p}) \bullet \mathbf{v}, \nabla g_2(\mathbf{p}) \bullet \mathbf{v}, \ldots, \nabla g_k(\mathbf{p}) \bullet \mathbf{v})).\end{aligned}$$

In other words,

$$\mathbf{X}'_\gamma(\mathbf{p}) = (\mathbf{p}, (D_{\mathbf{v}}g_1(\mathbf{p}), D_{\mathbf{v}}g_2(\mathbf{p}), \ldots, D_{\mathbf{v}}g_k(\mathbf{p}))). \tag{9.92}$$

Since the right-hand side of equation (9.92) does not use the curve $\gamma$ but only its tangent vector $\mathbf{v}$ at $\mathbf{p}$, we are done.

We can use equation (9.92) to relate the covariant derivative to ordinary directional derivatives in a cleaner way, but first another definition.

**Definition.** A *frame field* on a set $\mathbf{A} \subseteq \mathbf{R}^3$ is a triple of vector fields $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ defined over $\mathbf{A}$ with the property that, for every $\mathbf{p} \in \mathbf{A}$, $\mathbf{U}_1(\mathbf{p})$, $\mathbf{U}_2(\mathbf{p})$, and $\mathbf{U}_3(\mathbf{p})$ are an orthonormal basis of the tangent space $T_{\mathbf{p}}(\mathbf{R}^3)$. The frame field $(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3)$ on $\mathbf{R}^3$, where $\mathbf{E}_i(\mathbf{p}) = (\mathbf{p}, \mathbf{e}_i)$, is called the *standard frame field* of $\mathbf{R}^3$.

**9.16.2. Theorem.** Let $\mathbf{X}$ be an arbitrary vector field on $\mathbf{R}^3$ and suppose that

$$\mathbf{X} = \sum_{i=1}^{3} a_i \mathbf{E}_i$$

expresses the vector field $\mathbf{X}$ in terms of the standard frame field. If $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{R}^3)$, then

$$\nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p}) = \sum_{i=1}^{3} D_{\mathbf{v}}a_i(\mathbf{p})\mathbf{E}_i(\mathbf{p}).$$

**Proof.** This is clearly just a rewritten Equation (9.92). ($D_{\mathbf{v}}$ is just the ordinary directional derivative.)

The properties of the ordinary directional derivative carry over to the covariant derivative.

**9.16.3. Theorem.** Let $\mathbf{X}$ and $\mathbf{Y}$ be vector fields on $\mathbf{R}^3$ and let $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}(\mathbf{R}^3)$.

(1) $\nabla_{a\mathbf{v}+b\mathbf{w}}\mathbf{X}(\mathbf{p}) = a\nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p}) + b\nabla_{\mathbf{w}}\mathbf{X}(\mathbf{p})$ for all $a, b \in \mathbf{R}$.
(2) $\nabla_{\mathbf{v}}(a\mathbf{X} + b\mathbf{Y})(\mathbf{p}) = a\nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p}) + b\nabla_{\mathbf{v}}\mathbf{Y}(\mathbf{p})$ for all $a, b \in \mathbf{R}$.
(3) $\nabla_{\mathbf{v}}(f\mathbf{X})(\mathbf{p}) = D_{\mathbf{v}}f(\mathbf{p})\mathbf{X}(\mathbf{p}) + f(\mathbf{p})\nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p})$ for all differentiable functions $f: \mathbf{R}^3 \to \mathbf{R}$.
(4) $D_{\mathbf{v}}(\mathbf{X} \bullet \mathbf{Y})(\mathbf{p}) = \nabla_{\mathbf{v}}\mathbf{X}(\mathbf{p}) \bullet \mathbf{Y}(\mathbf{p}) + \mathbf{X}(\mathbf{p}) \bullet \nabla_{\mathbf{v}}\mathbf{Y}(\mathbf{p})$.

**Proof.** This is easy to prove directly but one can also use Theorem 9.16.2 and Proposition 4.3.18.

In the definitions above we defined the covariant derivative for vectors in a single fixed tangent space $T_{\mathbf{p}}(\mathbf{R}^3)$. It is easy to extend this definition.

**Definition.** Let $\mathbf{X}$ and $\mathbf{Y}$ be vector fields on $\mathbf{R}^3$. The *covariant derivative* of $\mathbf{X}$ with respect to $\mathbf{Y}$, denoted by $\nabla_{\mathbf{Y}}\mathbf{X}$, is the vector field defined by

$$\nabla_{\mathbf{Y}}\mathbf{X}(\mathbf{p}) = \nabla_{\mathbf{Y}(\mathbf{p})}\mathbf{X}(\mathbf{p}), \quad \text{for } \mathbf{p} \in \mathbf{R}^3.$$

We can consider $\nabla_{\mathbf{Y}}\mathbf{X}$ as a map

$$\begin{aligned} \mathrm{Vect}(\mathbf{R}^3) \times \mathrm{Vect}(\mathbf{R}^3) &\to \mathrm{Vect}(\mathbf{R}^3) \\ (\mathbf{X}, \mathbf{Y}) &\to \nabla_{\mathbf{Y}}\mathbf{X} \end{aligned} \tag{9.93}$$

that sends a pair of vector fields to another.

**9.16.4. Theorem.** Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{U}$, and $\mathbf{V}$ be vector fields on $\mathbf{R}^3$.

(1) $\nabla_{f\mathbf{U}+g\mathbf{V}}\mathbf{X} = f\nabla_{\mathbf{U}}\mathbf{X} + g\nabla_{\mathbf{V}}\mathbf{X}$ for all differentiable functions $f, g: \mathbf{R}^3 \to \mathbf{R}$.
(2) $\nabla_{\mathbf{U}}(a\mathbf{X} + b\mathbf{Y}) = a\nabla_{\mathbf{U}}\mathbf{X} + b\nabla_{\mathbf{U}}\mathbf{Y}$ for all $a, b \in \mathbf{R}$.
(3) $\nabla_{\mathbf{U}}(f\mathbf{X}) = (D_{\mathbf{U}}f)\mathbf{X} + f\nabla_{\mathbf{U}}\mathbf{X}$ for all differentiable functions $f: \mathbf{R}^3 \to \mathbf{R}$.
(4) $D_{\mathbf{U}}(\mathbf{X} \bullet \mathbf{Y}) = \nabla_{\mathbf{U}}\mathbf{X} \bullet \mathbf{Y} + \mathbf{X} \bullet \nabla_{\mathbf{U}}\mathbf{Y}$.

**Proof.** The proof is straightforward. One bit of notation needs explaining. The function $D_{\mathbf{U}}f: \mathbf{R}^3 \to \mathbf{R}$ in (3) is a vector field version of the directional derivative defined as follows: If $\mathbf{U}(\mathbf{p}) = (\mathbf{p}, \mathbf{u})$, then

$$D_{\mathbf{U}}f(\mathbf{p}) = (D_{\mathbf{u}}f)(\mathbf{p}).$$

So far we have just given lots of definitions with a few simple consequences, but we needed the terminology. As mentioned at the beginning of this section, we are

interested is the derivative of a vector field and its relation to the vector field. Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be a frame field on $\mathbf{R}^3$. Let $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{R}^3)$. Since the $\mathbf{U}_i$ define a basis at every point,

$$\nabla_{\mathbf{v}}\mathbf{U}_i(\mathbf{p}) = \sum_{j=1}^{3} c_{ij}\mathbf{U}_j(\mathbf{p}), \quad i = 1,2,3,$$

for some $c_{ij} \in \mathbf{R}$. It follows that

$$c_{ij} = \nabla_{\mathbf{v}}\mathbf{U}_i(\mathbf{p}) \bullet \mathbf{U}_j(\mathbf{p}).$$

Instead of a fixed vector $\mathbf{v}$ and a varying point $\mathbf{p}$, let us consider $\mathbf{p}$ fixed and $\mathbf{v}$ as varying. Define functions

$$\omega_{ij}(\mathbf{p}) : T_{\mathbf{p}}\left(\mathbf{R}^3\right) \to \mathbf{R}, \quad 1 \le i, j \le 3,$$

by

$$\omega_{ij}(\mathbf{p})(\mathbf{v}) = \nabla_{\mathbf{v}}\mathbf{U}_i(\mathbf{p}) \bullet \mathbf{U}_j(\mathbf{p}).$$

**Definition.** The function $\omega_{ij}$ are called the *connection forms* for the frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$.

The forms $\omega_{ij}$ have the following interpretation: If $\mathbf{v} = \mathbf{X}(\mathbf{p})$, then $\omega_{ij}(\mathbf{p})(\mathbf{v})$ describes the rate at which the vector field $\mathbf{U}_i$ rotates toward the vector $\mathbf{U}_j(\mathbf{p})$ as we move along a curve through $\mathbf{p}$ with tangent vector $\mathbf{v}$ at $\mathbf{p}$.

**9.16.5. Theorem.** The connection forms are differentiable 1-forms for the manifold $\mathbf{R}^3$ and satisfy $\omega_{ij} = -\omega_{ji}$.

**Proof.** To show that the $\omega_{ij}(\mathbf{p})$ are 1-forms one only needs to show that they are linear on each $T_{\mathbf{p}}(\mathbf{R}^3)$ and that is easy. For the second part, note that $\mathbf{U}_i \bullet \mathbf{U}_j = \delta_{ij}$, so that, using Theorem 9.16.3(4),

$$\begin{aligned} 0 &= D_{\mathbf{v}}(\delta_{ij})(\mathbf{p}) = D_{\mathbf{v}}(\mathbf{U}_i \bullet \mathbf{U}_j)(\mathbf{p}) \\ &= (\nabla_{\mathbf{v}}\mathbf{U}_i)(\mathbf{p}) \bullet \mathbf{U}_j(\mathbf{p}) + \mathbf{U}_i(\mathbf{p}) \bullet (\nabla_{\mathbf{v}}\mathbf{U}_j)(\mathbf{p}) \\ &= \omega_{ij}(\mathbf{p})(\mathbf{v}) + \omega_{ij}(\mathbf{p})(\mathbf{v}). \end{aligned}$$

The next theorem restates what we just showed.

**9.16.6. Theorem.** Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be a frame field on $\mathbf{R}^3$ and $\omega_{ij}$ its connection forms. If $\mathbf{X}$ is a vector field on $\mathbf{R}^3$, then

$$\nabla_{\mathbf{X}}\mathbf{U}_i(\mathbf{p}) = \sum_{j=1}^{3} \omega_{ij}(\mathbf{p})(\mathbf{X}(\mathbf{p}))\mathbf{U}_j(\mathbf{p}). \tag{9.94}$$

In fact, in expanded form we have

$$\begin{aligned}
\nabla_{\mathbf{X}}\mathbf{U}_1 &= \qquad\qquad\quad \omega_{12}(\mathbf{X})\mathbf{U}_2 + \omega_{13}(\mathbf{X})\mathbf{U}_3.\\
\nabla_{\mathbf{X}}\mathbf{U}_2 &= -\omega_{12}(\mathbf{X})\mathbf{U}_1 \qquad\qquad + \omega_{23}(\mathbf{X})\mathbf{U}_3.\\
\nabla_{\mathbf{X}}\mathbf{U}_3 &= -\omega_{13}(\mathbf{X})\mathbf{U}_1 - \omega_{23}(\mathbf{X})\mathbf{U}_2.
\end{aligned}$$

where $\omega_{ij}(\mathbf{X})(\mathbf{p})$ is an abbreviation for $\omega_{ij}(\mathbf{p})(\mathbf{X}(\mathbf{p}))$.

**Proof.** The fact that the equations can be expressed as shown using only the forms $\omega_{12}$, $\omega_{13}$, and $\omega_{23}$ follows from the fact that $\omega_{ij} = -\omega_{ji}$ for all i and j, which also implies that $\omega_{ii} = 0$.

Looking at the equations in Theorem 9.16.6 should remind the reader of the Serret-Frenet formulas. In fact, Exercise 9.16.1 shows that Theorem 9.16.6 is a generalization of these formulas. The Serret-Frenet formulas have no $\omega_{13}$ terms because of the special nature of the Frenet frames. Of course, the comparison might seem problematic because the Frenet frames are only defined along a curve and our connection forms in equations (9.94) were assumed to be defined at all points of $\mathbf{R}^3$. However, one can show that any frame field on a curve can always be extended to a tubular neighborhood of the curve so that equations (9.94) would have to hold as long as the vector field $\mathbf{X}$ is tangent to the curve because of the way that the directional derivative is defined (Lemma 9.16.1). Note that the derivatives in the Serret-Frenet formulas involve differentiating in a tangential direction.

**9.16.7. Example.** Let $\gamma(s)$ be a curve in $\mathbf{R}^2$ parameterized by arc-length. As usual, let $T(s) = \gamma'(s)$ and let $N(s)$ be the unit normal vector at $\gamma(s)$ so that $(T(s),N(s))$ induces the standard orientation on $\mathbf{R}^2$. Consider a frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{E}_3)$ defined in a neighborhood of the curve so that

$$\mathbf{U}_1(\gamma(s)) = T(s) \quad \text{and} \quad \mathbf{U}_2(\gamma(s)) = N(s).$$

Let $\mathbf{X} = \mathbf{U}_1$. Using the definition of the $\omega_{ij}$ and the notation of Theorem 9.16.6

$$\omega_{12}(\gamma(s))(T(s)) = \nabla_{T(s)}\mathbf{U}_1(\gamma(s)) \bullet \mathbf{U}_2(\gamma(s)) = \frac{d}{ds}T(s) \bullet N(s) = \boldsymbol{\kappa}_S N(s) \bullet N(s) = \boldsymbol{\kappa}_S.$$

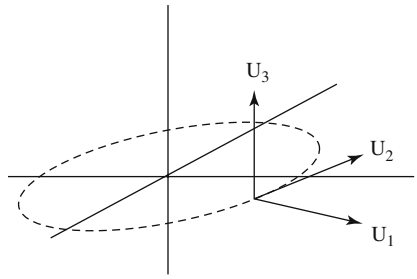
Furthermore, since the covariant derivative of $\mathbf{U}_1$ with respect to $S(s)$ lies in $\mathbf{R}^2$, we have

$$\omega_{13}(\gamma(s))(T(s)) = \nabla_{T(s)}\mathbf{U}_1(\gamma(s)) \bullet \mathbf{U}_3(\gamma(s)) = 0.$$

The form $\omega_{23}$ is zero for a similar reason. We see that using our special frame field and vector field $\mathbf{X}$ the covariant derivative equations (9.94) capture all of the geometry of the curve, namely, its curvature. In Exercise 9.16.1 you are asked to show an analogous fact for space curves.

It should be noted that all of our results above are local and rather than talking about vector fields and frame fields on $\mathbf{R}^3$ we could have developed the same results

**Figure 9.36.** The cylindrical frame field on $\mathbf{R}^3$.




with respect to some open subset (submanifold) of $\mathbf{R}^3$. In fact, any set for which one can make sense of the differentiability of functions would be adequate.

**9.16.8. Example.** To compute the connection forms for the *cylindrical frame field* $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ on $\mathbf{R}^3$ defined by

$$\begin{aligned}
\mathbf{U}_1(\mathbf{p}) &= (\cos\theta)\,\mathbf{E}_1(\mathbf{p}) + (\sin\theta)\,\mathbf{E}_2(\mathbf{p})\\
\mathbf{U}_2(\mathbf{p}) &= (-\sin\theta)\,\mathbf{E}_1(\mathbf{p}) + (\cos\theta)\,\mathbf{E}_2(\mathbf{p})\\
\mathbf{U}_3(\mathbf{p}) &= \mathbf{E}_3(\mathbf{p}),
\end{aligned}$$

where points are expressed in cylindrical coordinates $(r,\theta,z)$. See Figure 9.36.

**Solution.** Note that this field is not defined along the z-axis, but as long as we stay away from that axis everything that we did applies. Let $\mathbf{v}$ be a tangent vector at the point $\mathbf{p}$. By definition

$$\begin{aligned}
\nabla_{\mathbf{v}}\mathbf{U}_1(\mathbf{p}) &= (D_{\mathbf{v}}(\cos\theta))\,\mathbf{E}_1(\mathbf{p}) + (D_{\mathbf{v}}(\sin\theta))\,\mathbf{E}_2(\mathbf{p})\\
\nabla_{\mathbf{v}}\mathbf{U}_2(\mathbf{p}) &= (D_{\mathbf{v}}(-\sin\theta))\,\mathbf{E}_1(\mathbf{p}) + (D_{\mathbf{v}}(\cos\theta))\,\mathbf{E}_2(\mathbf{p})\\
\nabla_{\mathbf{v}}\mathbf{U}_3(\mathbf{p}) &= \mathbf{0}.
\end{aligned}$$

All that is left to do is to substitute into the formulas

$$\omega_{ij}(\mathbf{p})(\mathbf{v}) = \nabla_{\mathbf{v}}\mathbf{U}_i(\mathbf{p}) \bullet \mathbf{U}_j(\mathbf{p}).$$

For example,

$$\begin{aligned}
\omega_{12}(\mathbf{p})(\mathbf{v}) &= ((D_{\mathbf{v}}(\cos\theta))\,\mathbf{E}_1(\mathbf{p}) + (D_{\mathbf{v}}(\sin\theta))\,\mathbf{E}_2(\mathbf{p})) \bullet ((-\sin\theta)\,\mathbf{E}_1(\mathbf{p}) + (\cos\theta)\,\mathbf{E}_2(\mathbf{p}))\\
&= (-\sin\theta)\,D_{\mathbf{v}}(\cos\theta) + (\cos\theta)\,D_{\mathbf{v}}(\sin\theta).
\end{aligned}$$

Using the relationships

$$\begin{aligned}
D_{\mathbf{v}}(\cos\theta) &= (-\sin\theta)\,d\theta(\mathbf{v})\\
D_{\mathbf{v}}(\sin\theta) &= (\cos\theta)\,d\theta(\mathbf{v})
\end{aligned}$$

between directional derivatives and differentials (equation (4.31b) in Chapter 4), it follows that $\omega_{12} = d\theta$. This implies that $\omega_{21} = -d\theta$. It is easy to see that all other $\omega$'s are zero.

To understand the motion of frame fields better, one introduces dual forms.

**Definition.** Given a frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ on $\mathbf{R}^3$, define the *dual forms* $\theta_i(\mathbf{p}) \in T_{\mathbf{p}}(\mathbf{R}^3)^*$ by

$$\theta_i(\mathbf{p})(\mathbf{v}) = \mathbf{v} \bullet \mathbf{U}_i(\mathbf{p}), \quad \mathbf{v} \in T_{\mathbf{p}}(\mathbf{R}^3).$$

As usual, $\theta_i(\mathbf{p})$ will be abbreviated to $\theta_i$ unless the point $\mathbf{p}$ needs to be specified explicitly.

**9.16.9. Theorem.** (The Cartan structural equations) Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be a frame field on $\mathbf{R}^3$. Let $\omega_{ij}$ and $\theta_i$ be the connection and dual forms, respectively, of this frame field. Then

(1) (The first structural equations) $$d\theta_i = \sum_{j=1}^{3} \omega_{ij} \wedge \theta_j$$

(2) (The second structural equations) $$d\omega_{ij} = \sum_{k=1}^{3} \omega_{ik} \wedge \omega_{kj}$$

**Proof.** See [ONei66].

Next, we move on to surfaces. Just like in the case of curves in $\mathbf{R}^3$, the geometry of a surface can be deduced by analyzing special frame fields on them.

Let $\mathbf{S}$ be a surface in $\mathbf{R}^3$.

**Definition.** A frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ on $\mathbf{S}$ is called an *adapted frame field* on $\mathbf{S}$ if $\mathbf{U}_3$ is a normal vector field for $\mathbf{S}$.

Again, everything we do will be local in nature, so that we do not need adapted frame fields to be defined on all of $\mathbf{S}$, only on an open subset that is relevant at the time. Because adapted frame fields are somewhat special, it is useful to make clear when they exist.

**9.16.10. Proposition.** A surface $\mathbf{S}$ admits an adapted frame field if and only if it is orientable and has a nonzero (tangent) vector field.

**Proof.** Easy.

**9.16.11. Example.** The cylinder defined by the equation

$$x^2 + y^2 = r^2$$

admits the adapted frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ defined by

$$\mathbf{U}_1(x,y,z) = \mathbf{U}_2(x,y,z) \times \mathbf{U}_3(x,y,z) = \frac{1}{r}(-y\mathbf{E}_1(x,y,z) + x\mathbf{E}_2(x,y,z))$$
$$\mathbf{U}_2(x,y,z) = \mathbf{E}_3(x,y,z)$$
$$\mathbf{U}_3(x,y,z) = \frac{1}{r}(x\mathbf{E}_1(x,y,z) + y\mathbf{E}_2(x,y,z)).$$

**9.16.12. Example.** Consider the sphere defined by the equation

$$x^2 + y^2 + z^2 = r^2.$$

Because the sphere does not have any nonzero vector field (Corollary 8.5.6) there is no adapted frame field defined over all of it. On the other hand we can, for example, define an adapted frame field $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ over the sphere minus the north and south pole by defining

$$\mathbf{U}_1(x,y,z) = \frac{1}{\sqrt{x^2+y^2}}(-y,x,0) = (-y\mathbf{E}_1(x,y,z) + x\mathbf{E}_2(x,y,z))$$
$$\mathbf{U}_2(x,y,z) = \mathbf{U}_3(x,y,z) \times \mathbf{U}_1(x,y,z)$$
$$\mathbf{U}_3(x,y,z) = \frac{1}{r}(x\mathbf{E}_1(z,y,z) + y\mathbf{E}_2(x,y,z) + z\mathbf{E}_3(x,\mathbf{y},z)).$$

If the sphere were the earth, then the vector $\mathbf{U}_1$ would point due "east." See Figure 9.37.

Next, assume that $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ is an adapted frame field on a surface $\mathbf{S}$. Although the frame field is only defined at points $\mathbf{p}$ in the surface $\mathbf{S}$, as long as we stick to vectors $\mathbf{v}$ in $T_{\mathbf{p}}(\mathbf{S})$, the connection equation

$$\nabla_{\mathbf{v}}\mathbf{U}_i = \sum_{j=1}^{3} \omega_{ij}(\mathbf{p})(\mathbf{v})\mathbf{U}_j(\mathbf{p}).$$

will still be defined. This follows again because of the way that the directional derivative is defined—we differentiate in the tangent direction of a curve in the surface

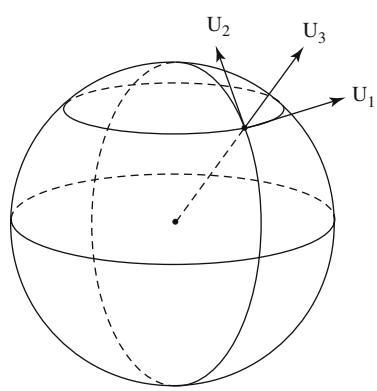



**Figure 9.37.** The adaptive frame field on $\mathbf{S}^2$ in Example 9.16.12.

through the point. We can relate the frame field to the shape operator for **S** since $\mathbf{U}_3$ is a normal vector.

**9.16.13. Theorem.** Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be an adapted frame field for a surface **S**. If S is the shape operator for **S** with respect to the normal vector field $\mathbf{U}_3$, then for every $\mathbf{v} \in T_{\mathbf{p}}(\mathbf{S})$ we have

$$S(\mathbf{v}) = \omega_{13}(\mathbf{v})\mathbf{U}_1(\mathbf{p}) + \omega_{23}(\mathbf{v})\mathbf{U}_2(\mathbf{p}).$$

**Proof.** This follows easily from the definition of S, namely, $S(\mathbf{v}) = -\nabla_{\mathbf{v}}\mathbf{U}_3$.

Define the dual forms $\theta_i$ as before. The form $\theta_3$ will be zero, so that we shall ignore it.

**9.16.14. Theorem.** (The Cartan structural equations) Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be an adapted frame field for a surface **S**. Let $\omega_{ij}$ and $\theta_i$ be the connection and dual forms, respectively, of this frame field. Then

| | |
|---|---|
| (1) (The first structural equations) | $d\theta_1 = \omega_{12} \wedge \theta_2$ |
| | $d\theta_2 = \omega_{21} \wedge \theta_1$ |
| (2) (The symmetry equation) | $\omega_{31} \wedge \theta_1 + \omega_{32} \wedge \theta_2 = 0$ |
| (3) (The Gauss equation) | $d\omega_{12} = \omega_{13} \wedge \omega_{32}$ |
| (4) (The Codazzi equations) | $d\omega_{13} = \omega_{12} \wedge \omega_{23}$ |
| | $d\omega_{23} = \omega_{21} \wedge \omega_{13}$ |

**Proof.** See [ONei66].

All the geometric invariants of surfaces can be deduced from Theorem 9.16.14. There is no space to prove anything here, but it is instructive to indicate how Gauss curvature K and the mean curvature H come into the picture.

**9.16.15. Lemma.**

(1) $\omega_{13} \wedge \omega_{23} = K\, \theta_1 \wedge \theta_2$.
(2) $\omega_{13} \wedge \theta_2 + \theta_1 \wedge \omega_{23} = 2H\, \theta_1 \wedge \theta_2$.

**Proof.** See [ONei66].

**9.16.16. Corollary.** (The second structural equation) $d\omega_{12} = -K\, \theta_1 \wedge \theta_2$.

**Proof.** This follows from Theorem 9.16.14(3) and Lemma 9.16.15(1).

Finally, frame fields can also be defined for abstract surfaces (manifolds). There will not be any normal vector because that would not make sense, but one can talk about frame fields $(\mathbf{U}_1,\mathbf{U}_2)$, where $\mathbf{U}_1$ and $\mathbf{U}_2$ form an orthonormal basis in each tangent space. There is no problem about defining the corresponding differential forms $\omega_{ij}$ and $\theta_i$. The first structural equations would again hold. Since one has no

normal vector, one cannot talk about any Gauss map nor define the Gauss curvature directly as before, but now one would be able to prove that the second structural equation holds for some function K and this function would then be defined to be the *Gauss curvature*. Now, the forms $\omega_{ij}$ and $\theta_i$ depend on the frame field, but the function K turns out to be independent of it. In the end, one would therefore again have the first and second structural equations hold. Hence one would get the same geometric consequences as before. The Christoffel symbols are also embedded in all this. See [Spiv70b].

## 9.17 Where to Next?

In this chapter we have taken the first steps in learning about some of the beautiful geometric results as seen through the eyes of a differential geometer. From a historical perspective, our presentation may have been more modern but we did not learn much more than what was already known at the time of Gauss. We have basically looked at results that can be deduced from the first and second fundamental forms. We have seen how a "simple" map like the Gauss map can contain within it a wealth of information.

Euler's formula about the Euler characteristic showed that some properties of spaces did not depend on their metric. Gauss showed that Gauss curvature depended only on arc-length in the surface. Another important early result is the Gauss-Bonnet theorem below, which generalizes the theorem in the plane that the sum of angles of a triangle is $\pi$. It shows a connection between the metric invariant, Gauss curvature, and a topological invariant, the Euler characteristic, and generalizes the relationship between the total curvature of a closed curve and its degree (Theorem 9.3.13 and Corollary 9.3.14).

**9.17.1. Theorem.** (Gauss-Bonnet) Let **S** be a compact, closed, orientable surface. Then

$$\iint_{\mathbf{S}} K = 2\pi\chi(\mathbf{S}), \tag{9.95}$$

where K is the Gauss curvature function on **S** and $\chi(\mathbf{S})$ is the Euler characteristic of **S**. (The integral in this equation is called the *total Gauss curvature* of **S**.)

**Proof.** See one of the references for differential geometry.

One of the interesting consequences of this theorem is what it says about the Gauss curvature function. Since the right-hand side of Equation (9.95) is a topological invariant, no matter how we deform a surface, there are constraints as to how the Gauss curvature distributes itself. It is not a totally arbitrary function. Here is an application of the Gauss-Bonnet theorem.

**9.17.2. Corollary.** A compact, closed, orientable surface whose Gauss curvature is positive everywhere must be homeomorphic to a sphere.

**Proof.** Equation (9.95) implies that the Euler characteristic of the surface is positive and the classification theorem for surface tells us that the sphere is only such surface.

There are many other beautiful theorems showing that the topology of a surface is influenced by its curvature.

It was Riemann who took differential geometry to the next plateau. His inaugural lecture at Göttingen in 1854 has been considered to be one of the most influential in the field. See [Spiv70b] for an excellent discussion of what Riemann said. He moved differential geometry from surfaces in $\mathbf{R}^3$ to n-dimensional manifolds that are endowed with an intrinsic metric (now called a Riemannian metric), which is defined independent of approximations by straight lines (lengths of curves in calculus were traditionally defined in terms approximating polygonal lines with the length of a straight line segment as basis). Starting with an arbitrary metric

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}(\mathbf{x})dx_i dx_j \tag{9.96}$$

Riemann asked the question "when is this metric isometric to a given one?" For example, perhaps a change of coordinates would turn the metric into

$$ds^2 = \sum_{i=1}^{n} dx_i^2,$$

that is, we have a "flat" space, one that is isometric to $\mathbf{R}^n$. Now, using the Taylor expansion for $g_{ij}(\mathbf{x})$ we can rewrite equation (9.96) in the form

$$ds^2 = \sum_{i,j=1}^{n}\left(\delta_{ij} + \sum_{r,t=1}^{n} \frac{\partial g_{ij}}{\partial x_r \partial x_t}(0)x_r x_t + \ldots\right)dx_i dx_j. \tag{9.97}$$

By analyzing the dominating expression of second partials, Riemann was led to the curvature tensor, which is an appropriate function of these partials. The modern way to approach this subject is by means of what is called a connection. This is where Section 9.16 was leading. Although it will take a string of definitions and we will give no proofs, we feel that it is worthwhile to give the reader an inkling of what one has to do.

Let $\mathbf{M}^n$ be an arbitrary Riemannian manifold. To keep things concrete, the user may assume that the manifold is a submanifold of some $\mathbf{R}^m$ and a vector field simply assigns to every point of $\mathbf{M}$ a vector in $\mathbf{R}^m$ that is tangent to the manifold.

**Notation.** If $\eta(t)$ is a curve in $\mathbf{M}$, then $\mathrm{Vect}(\mathbf{M},\eta)$ will denote the set of vector fields $\mathbf{X}(t)$ defined along $\eta$. The zero vector field in $\mathrm{Vect}(\mathbf{M})$ or $\mathrm{Vect}(\mathbf{M},\eta)$ will be denoted by $\mathbf{0}$.

We want a general notion of a covariant derivative of one vector field with respect to another one. It should be something that generalizes the covariant derivative of a

vector field along a curve that was defined in Section 9.10 and at the same time specialize to that one.

**Definition.** A map

$$\begin{aligned} \mathrm{Vect}(\mathbf{M}) \times \mathrm{Vect}(\mathbf{M}) &\to \mathrm{Vect}(\mathbf{M}) \\ (\mathbf{X}, \mathbf{Y}) &\to \nabla_{\mathbf{X}} \mathbf{Y} \end{aligned}$$

is called a *connection* or *covariant derivative* on $\mathbf{M}$ if it satisfies the following properties. Let $f \in C^{\infty}(\mathbf{M})$.

(1) $\nabla_{\mathbf{X}}\mathbf{Y}$ is a bilinear function over $\mathbf{R}$ in $\mathbf{X}$ and $\mathbf{Y}$.
(2) $\nabla_{f\mathbf{X}}\mathbf{Y} = f\nabla_{\mathbf{X}}\mathbf{Y}$.
(3) $\nabla_{\mathbf{X}} f\mathbf{Y} = (\mathbf{X}f)\mathbf{Y} + f\nabla_{\mathbf{X}}\mathbf{Y}$, where $\mathbf{X}f(\mathbf{p})$ is the directional derivative of f in the direction $\mathbf{X}$ at $\mathbf{p}$.

The vector field $\nabla_{\mathbf{X}}\mathbf{Y}$ is called the *covariant derivative* of the vector field $\mathbf{Y}$ with respect to the vector field $\mathbf{X}$.

Let

$$\Phi : \mathbf{U} \to \mathbf{M}$$

define a coordinate neighborhood for a point $\mathbf{p} \in \mathbf{M}$. We shall use $\partial/\partial u_i$ (same as $\Phi_i$) to denote the standard basis vectors of the tangent spaces. Therefore in terms of this basis,

$$\nabla_{\partial/\partial u_i} \partial/\partial u_j = \sum_{k=1}^{n} \Gamma_{ij}^{k} \, \partial/\partial u_k .$$

The functions $\Gamma_{ij}^{k}$ determine the connection completely and are called the *Christoffel symbols* of the connection. In general, they are not the same functions that we defined in Section 9.9 without some additional hypotheses. Given a connection on the manifold $\mathbf{M}$ and a curve $\eta(t)$ in $\mathbf{M}$ we can define a covariant derivative along curves.

**Definition.** Given a connection $\nabla_{\mathbf{X}}\mathbf{Y}$ on $\mathbf{M}$, a *compatible covariant derivative along a curve* $\eta$: $[a,b] \to \mathbf{M}$ is a map

$$\frac{D}{dt} : \mathrm{Vect}(\mathbf{M}, \eta) \to \mathrm{Vect}(\mathbf{M}, \eta)$$

that sends a vector field $\mathbf{X}(t)$ along $\eta$ into a new vector field $(D\mathbf{X}/dt)(t)$ along $\eta$ satisfying the following properties. Assume that $\mathbf{X}(t)$, $\mathbf{Y}(t) \in \mathrm{Vect}(\mathbf{M},\eta)$ and that $f:[a,b] \to \mathbf{R}$ is a differentiable function.

(1) $\dfrac{D}{dt}(\mathbf{X} + \mathbf{Y}) = \dfrac{D\mathbf{X}}{dt} + \dfrac{D\mathbf{Y}}{dt}$.
(2) $\dfrac{D}{dt}(f\mathbf{X}) = \dfrac{df}{dt}\mathbf{X} + f\dfrac{D\mathbf{X}}{dt}$.

(3) Given any $\mathbf{Z} \in \mathrm{Vect}(\mathbf{M})$ that agrees with $\mathbf{X}$ on $\eta$ and any $\mathbf{T} \in \mathrm{Vect}(\mathbf{M})$ that agrees with $\eta'(t)$ on $\eta$, then

$$\frac{D\mathbf{X}}{dt}(t) = (\nabla_{\mathbf{T}}\mathbf{Z})(\eta(t)).$$

(A vector field $\mathbf{A} \in \mathrm{Vect}(\mathbf{M})$ *agrees* with a vector field $\mathbf{B} \in \mathrm{Vect}(\mathbf{M},\eta)$ on $\eta$ if $\mathbf{A}(\eta(t)) = \mathbf{B}(t)$.)

One can show that every connection defines a unique covariant derivative along curves. This is an easy consequence of the axioms. Specifically, let

$$\mathbf{X} = \sum_{i=1}^{n} a_i \frac{\partial}{\partial u_i} \quad \text{and} \quad \gamma = \Phi^{-1}\circ\eta : [a,b] \to \mathbf{R}^n,$$

where $a_i : [a,b] \to \mathbf{R}$. Then

$$\frac{D\mathbf{X}}{dt} = \sum_{k=1}^{n}\left[\frac{da_k}{dt} + \sum_{i,j=1}^{n}\frac{d\gamma_i}{dt}\Gamma_{ij}^{k}a_j\right]\frac{\partial}{\partial u_k}.$$

**Definition.** A vector field $\mathbf{X}(t)$ along a curve $\eta$ is called a *parallel vector field* along $\eta$ if $D\mathbf{X}/dt = \mathbf{0}$.

One can show just like in Section 9.10 that a vector at the start point of a curve defines a unique parallel vector field along the curve.

**Definition.** A connection on $\mathbf{M}$ is said to be *compatible with the metric* on $\mathbf{M}$ if any pair of parallel vector fields along a curve $\eta(t)$ have a constant inner product, that is, if $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ are parallel vector fields along $\eta$, then $<\mathbf{X}(t),\mathbf{Y}(t)>$ is constant.

If one has a connection compatible with the metric, then the covariant derivative along a curve satisfies

$$\frac{d}{dt}<\mathbf{X},\mathbf{Y}> = <\frac{D\mathbf{X}}{dt},\mathbf{Y}> + <\mathbf{X},\frac{D\mathbf{Y}}{dt}>.$$

**Definition.** A connection $\nabla_{\mathbf{X}}\mathbf{Y}$ on $\mathbf{M}$ is said to be *symmetric* or *torsion-free* if it satisfies

$$\nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} = [\mathbf{X},\mathbf{Y}],$$

where $[\mathbf{X},\mathbf{Y}]$ is the vector field defined by

$$[\mathbf{X},\mathbf{Y}]_{\mathbf{p}}f = \mathbf{X}_{\mathbf{p}}(\mathbf{Y}f) - \mathbf{Y}_{\mathbf{p}}(\mathbf{X}f)$$

for $\mathbf{p} \in \mathbf{M}$ and $f \in C^{\infty}(\mathbf{M})$ called the *Lie bracket* of $\mathbf{X}$ and $\mathbf{Y}$.

**9.17.3. Theorem.** (The Fundamental Lemma of Riemannian Geometry) A Riemannian manifold $\mathbf{M}^n$ possesses a unique symmetric connection that is compatible with its metric. This connection is called the *Levi-Civita connection* for the metric.

**Proof.** See [Spiv75] or [MilP77]. The Christoffel symbols defined above are the same as the Christoffel symbols defined in Section 9.9 for this connection.

**Definition.** Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z} \in \mathrm{Vect}(\mathbf{M})$. The vector field $\mathrm{R}(\mathbf{X},\mathbf{Y})\mathbf{Z}$ defined by

$$\mathrm{R}(\mathbf{X},\mathbf{Y})\mathbf{Z} = -\nabla_{\mathbf{X}}(\nabla_{\mathbf{Y}}\mathbf{Z}) + \nabla_{\mathbf{Y}}(\nabla_{\mathbf{X}}\mathbf{Z}) + \nabla_{[\mathbf{X},\mathbf{Y}]}\mathrm{Z}$$

is called the *curvature tensor* of the connection.

The key to Riemannian geometry is an analysis of the curvature tensor. (Tensors were briefly alluded to in Section 4.9 and are a whole other large topic in differential geometry and physics.) For example,

**9.17.4. Theorem.** If $\mathbf{M}^n$ is a Riemannian manifold, then the curvature tensor R for the Levi-Civita connection on $\mathbf{M}^n$ is 0 if and only if $\mathbf{M}^n$ is locally isometric to $\mathbf{R}^n$ with its usual Riemannian metric.

**Proof.** See [Spiv75].

It must be pointed out that Theorem 9.17.4 is a local theorem because there are manifolds that are **locally** isometric to $\mathbf{R}^n$ but are not homeomorphic to $\mathbf{R}^n$. See [Spiv75].

Here is another way that one can see the geometry that is embedded in the concept of the curvature tensor. The latter is really defined by lower-dimensional concepts that are easier to understand. Let $\mathbf{M}^n$ be a Riemannian manifold with curvature tensor R. Let $\mathbf{p} \in \mathbf{M}$ and let $\mathbf{V}$ be a two-dimensional subspace of the tangent space $\mathrm{T}_{\mathbf{p}}(\mathbf{M})$. It turns out that the geodesics in a neighborhood of $\mathbf{p}$ that pass through $\mathbf{p}$ and are tangent to $\mathbf{V}$ at $\mathbf{p}$ actually define a piece of a surface $\mathbf{S}$ through $\mathbf{p}$ that is uniquely defined by $\mathbf{V}$. The surface $\mathbf{S}$ has a Riemannian structure induced by the structure on $\mathbf{M}$.

**Definition.** The Gauss curvature of $\mathbf{S}$ at $\mathbf{p}$ is called the *sectional curvature* of $\mathbf{M}$ at $\mathbf{p}$ along $\mathbf{V}$ and is denoted by $\mathrm{K}(\mathbf{p},\mathbf{V})$.

**9.17.5. Theorem.** With the notation and hypotheses leading up to the definition of sectional curvature, let $\mathbf{X}$ and $\mathbf{Y}$ be vectors in $\mathrm{T}_{\mathbf{p}}(\mathbf{M})$ that span $\mathbf{V}$. Then

$$\mathrm{K}(\mathbf{p},\mathbf{V}) = \frac{<\mathrm{R}(\mathbf{X},\mathbf{Y})\mathbf{Y},\mathbf{X}>}{\mathrm{A}},$$

where A is the area of the parallelogram spanned by $\mathbf{X}$ and $\mathbf{Y}$.

**Proof.** See [Spiv75].

Theorem 9.17.5 says that the curvature tensor at a point is defined by the Gauss curvatures of the surfaces through the point.

Returning to the subject of geodesics, we now have all the tools needed to define this concept for arbitrary Riemannian manifolds. Of course, there never was a problem with a generalization if we base a definition on the property that the curve is locally of shortest length. Let us quickly look over the definitions we gave in Section 9.10. The first definition we gave for geodesics in surfaces does not easily generalize. The second definition does easily generalize to curves in an n-dimensional manifold $\mathbf{M}^n$ in $\mathbf{R}^{n+1}$ because we again have a normal vector. The fourth does also, because generalizing the definition of parallel vector fields is no problem in that case. See [Thor79]. For arbitrary manifolds, where we have no normal vector, it is the definition of a geodesic based on parallel vector fields, which is generalized. The key to this generalization is a definition of the covariant derivative of a vector field along a curve that we have now.

**Definition.** A curve $\eta(t)$ in a Riemannian manifold $\mathbf{M}^n$ is called a *geodesic* if the covariant derivative of its tangent vector field vanishes, that is, $(D/dt)\eta' = 0$.

With this definition, it is then straighforward to show that geodesics have the same properties that they had for surfaces. In particular, they are defined by second order differential equations and there is a unique geodesic that starts at a point with a given tangent vector there. There is also an exponential map that maps a neighborhood of $\mathbf{0}$ in the tangent space at a point diffeomorphically onto a neighborhood of the point in the manifold.

Finally, it is interesting to point out that the geometric invariants for curves and surfaces that were discussed here can also be derived from a more analytical point of view. Look back to Equation (9.97) and the comments following it. As a simpler example, suppose that we want to study a curve $\mathbf{C}$ at a point $\mathbf{p}$. There is no loss of generality (by moving the curve with a rigid motion) in assuming the $\mathbf{p}$ is the origin and the tangent line is just the x-axis. We may also assume that the curve is a graph of a function f(x) in a neighborhood of $\mathbf{p}$. In that case, the Taylor expansion for f at the origin is

$$f(x) = (1/2)cx^2 + \dots.$$

It turns out that the constant c is just the curvature of the curve. A similar analysis can be made for a surface $\mathbf{S}$. In that case we may also assume that the point of interest on the surface is the origin, the tangent plane is just the xy-plane, and the surface is the graph of a function f(x,y). The two-dimensional Taylor expansion for f is now

$$f(x,y) = (1/2)\left(ax^2 + 2bxy + cy^2\right) + \dots,$$

where $a = f_{xx}(0,0)$, etc. The principal normal curvatures are just the eigenvalues of the quadratic form $ax^2 + 2bxy + cy^2$, and so on.

## 9.18 Summary of Curve Formulas

Length of a curve $\gamma:[a,b] \to \mathbf{R}^n$: $\int_a^b |\gamma'|$

Arc-length parameterization $\gamma(s)$: $|\gamma'(s)| = 1$ (this equation defines it)
$\gamma'(s) \bullet \gamma''(s) = 0$

In general:
$$\frac{d}{ds} = \frac{1}{|\gamma'(t)|}\frac{d}{dt}$$
$T(s) = \gamma'(s)$

Curve $\gamma$ in $\mathbf{R}^2$:

Principal normal $N(s)$: $|N(s)| = 1$
$(T(s),N(s))$ determines the standard orientation of $\mathbf{R}^2$
$N(s) = (-T_2(s),T_1(s))$

Signed curvature $\kappa_S(s)$: $T'(s) = \kappa_S(s)N(s)$

Curvature $\kappa(s)$: $\kappa(s) = |\kappa_S(s)| = |T'(s)|$

$$\kappa_S(s) = \det\begin{pmatrix} T(s) \\ T'(s) \end{pmatrix}$$

$$\kappa_S(t) = \frac{\gamma_1'(t)\gamma_2''(t) - \gamma_2'(t)\gamma_1''(t)}{\left(\gamma_1'^2(t) + \gamma_2'^2(t)\right)^{3/2}} = \frac{1}{|\gamma'(t)|^3}\det\begin{pmatrix} \gamma'(t) \\ \gamma''(t) \end{pmatrix}$$

If $n(t) = \frac{1}{|\gamma'(t)|}(-\gamma_2'(t), \gamma_1'(t))$, then $n'(t) = \kappa_S(t)\gamma'(t)$.

Total curvature $\kappa_T$ of $\gamma$: $\frac{d}{ds}\kappa_T(s) = \kappa_S(s)$

Curve $\gamma$ in $\mathbf{R}^3$:

Curvature vector: $K(s) = \gamma''(s)$

Curvature: $\kappa(s) = |K(s)|$
$$\kappa(t) = \frac{|\gamma'(t) \times \gamma''(t)|}{|\gamma'(t)|^3}$$

Inflection point: Point on curve where curvature is zero.

Principal normal $N(s)$: $T'(s) = \kappa(s)N(s)$

Binormal $B(s)$: $B(s) = T(s) \times N(s)$

Torsion $\tau(s)$: $B'(s) = -\tau(s)N(s)$

$$\tau(s) = \frac{1}{\kappa^2(s)}(\gamma'(s) \times \gamma''(s)) \bullet \gamma'''(s) = \frac{1}{\kappa^2(s)} \det\begin{pmatrix} \gamma'(s) \\ \gamma''(s) \\ \gamma'''(s) \end{pmatrix}$$

$$\tau(t) = \frac{(\gamma'(t) \times \gamma''(t)) \bullet \gamma'''(t)}{|\gamma'(t) \times \gamma''(t)|^2} = \frac{1}{|\gamma'(t) \times \gamma''(t)|^2} \det\begin{pmatrix} \gamma'(t) \\ \gamma''(t) \\ \gamma'''(t) \end{pmatrix}$$

The Serret-Frenet Formulas:

$$\begin{aligned} T' &= \kappa N \\ N' &= -\kappa T + \tau B \\ B' &= -\tau N \end{aligned}$$

Curve $\gamma$ in $\mathbf{R}^n$:

$$\begin{aligned} u_1' &= \kappa_1 u_2 \\ u_2' &= -\kappa_1 u_1 + \kappa_2 u_3 \\ u_3' &= -\kappa_2 u_2 + \kappa_3 u_4 \\ &\vdots \\ u_{n-1}' &= -\kappa_{n-2} u_{n-2} + \kappa_{n-1} u_n \\ u_n' &= -\kappa_{n-1} u_{n-1} \end{aligned}$$

Involute of curve $p(s)$: $p^*(s) = p(s) + (c - s)T(s)$, where c is a constant

Curvature of involute $p^*(s)$: $\kappa^* = \dfrac{\kappa^2 + \tau^2}{(c-s)^2 \kappa^2}$

Evolute of curve $p(s)$: $q(s) = p(s) + \dfrac{1}{\kappa(s)} N(s) + \dfrac{1}{\kappa(s)} \cot\left(\int_0^s \tau + c\right) B(s)$,
for some constant c.

Evolute of planar curve $p(s)$: $q(s) = p(s) + \dfrac{1}{\kappa(s)} N(s) + \dfrac{c}{\kappa(s)} B(s)$,
for some constant c

Plane evolute of planar curve $p(s)$: $q(s) = p(s) + \dfrac{1}{\kappa(s)} N(s)$

Parallel or offset curves:

$$\begin{aligned} p_d(t) &= p(t) + dn(t) \\ p_d'(t) &= p'(t)\,(1 + \kappa(t)d) \\ \kappa_d(t) &= \frac{\kappa(t)}{1 + \kappa(t)d} \end{aligned}$$

## 9.19 Summary of Surface Formulas

In the formulas below $\Phi(u,v)$ is a parameterization of a surface $\mathbf{S}$ in $\mathbf{R}^3$ and

$$\Phi_u = \frac{\partial \Phi}{\partial u} \quad \text{and} \quad \Phi_v = \frac{\partial \Phi}{\partial v}.$$

$\gamma(t)$ will be a regular curve on $\mathbf{S}$ and assumed to be expressed in the form $\gamma(t) = \Phi(\mu(t))$, where $\mu(t) = (u(t),v(t))$ is a curve in $\mathbf{R}^2$.

First fundamental form: A quadratic form $Q_I$ defined on the tangent space $T_{\mathbf{p}}(\mathbf{S})$ by

$$Q_I(\mathbf{v}) = \mathbf{v} \bullet \mathbf{v}$$

Metric coefficients:

For first fundamental form: $Q_I(\gamma') = E(u')^2 + 2F(u')(v') + G(v')^2$

$$E = \Phi_u \bullet \Phi_u,\ F = \Phi_u \bullet \Phi_v,\ G = \Phi_v \bullet \Phi_v$$

$$|\Phi_u \times \Phi_v|^2 = EG - F^2$$

For a general parameterization $\Phi(u_1,u_2, \ldots ,u_n)$: $g_{ij} = \dfrac{\partial \Phi}{\partial u_i} \bullet \dfrac{\partial \Phi}{\partial u_j}$

Length: $\text{Length}(\gamma) = \int_0^t \sqrt{Eu'^2 + 2Fu'v' + Gv'^2}$

Angle $\alpha$ between $\Phi_u$ and $\Phi_v$: $\cos\alpha = \dfrac{\Phi_u \bullet \Phi_v}{|\Phi_u||\Phi_v|} = \dfrac{F}{\sqrt{EG}}$

Area of $\mathbf{S}$ with $\mathbf{U}$ the domain of $\Phi$: $\int_{\mathbf{U}} \sqrt{EG - F^2}$

Volume of a parameterization $\Phi$ with domain $\mathbf{U}(u_1,u_i, \ldots ,u_n)$ for a manifold $\mathbf{M}^n$ in $\mathbf{R}^{n+1}$:

$$\begin{aligned} V(\Phi) &= \int_{\mathbf{U}} \sqrt{\det(g_{ij})} \\ &= \int_{\mathbf{U}} \sqrt{g}, \quad \text{where } g = \det(g_{ij}) \end{aligned}$$

Gauss map: $\mathbf{n}: \mathbf{S} \to \mathbf{S}^2$, where $\mathbf{n}(\mathbf{p})$ is the unit normal at $\mathbf{p}$ for the oriented surface $\mathbf{S}$

Second fundamental form: A quadratic form $Q_{II}$ defined on the tangent space $T_{\mathbf{p}}(\mathbf{S})$ by

$$Q_{II}(\mathbf{v}) = -D\mathbf{n}(\mathbf{p})(\mathbf{v}) \bullet \mathbf{v}$$

Coefficients of second fundamental form: $Q_{II}(v) = L(u')^2 + 2M(u')(v') + N(v')^2$

$$\begin{aligned} L &= -\mathbf{n}_u \bullet \Phi_u = \mathbf{n} \bullet \Phi_{uu}, \\ M &= -\mathbf{n}_v \bullet \Phi_u = \mathbf{n} \bullet \Phi_{uv} = \mathbf{n} \bullet \Phi_{vu} = -\mathbf{n}_u \bullet \Phi_v \\ N &= -\mathbf{n}_v \bullet \Phi_v = \mathbf{n} \bullet \Phi_{vv}. \end{aligned}$$

Shape operator: $S_{\mathbf{p}} = -D\mathbf{n}(\mathbf{p}) : T_{\mathbf{p}}(\mathbf{S}) \to T_{\mathbf{p}}(\mathbf{S})$

Normal curvature of $\gamma$ at $\mathbf{p}$ in $\mathbf{S}$: $k_{n,\gamma}(\mathbf{p}) = \boldsymbol{\kappa}\mathbf{N} \bullet \mathbf{n}(\mathbf{p})$,
where $\mathbf{N}$ is the principal normal and $\boldsymbol{\kappa}$ is the curvature to $\gamma$ at $\mathbf{p}$.

Choose orthonormal basis $(\mathbf{u}_1, \mathbf{u}_2)$ for $T_{\mathbf{p}}(\mathbf{S})$ so that

$$\begin{aligned} D\mathbf{n}(\mathbf{p})(\mathbf{u}_1) &= -\boldsymbol{\kappa}_1\mathbf{u}_1 \qquad (S_{\mathbf{p}}(\mathbf{u}_1) = \boldsymbol{\kappa}_1\mathbf{u}_1) \\ D\mathbf{n}(\mathbf{p})(\mathbf{u}_2) &= -\boldsymbol{\kappa}_2\mathbf{u}_2 \qquad (S_{\mathbf{p}}(\mathbf{u}_2) = \boldsymbol{\kappa}_2\mathbf{u}_2). \end{aligned}$$

and assume that $\boldsymbol{\kappa}_1 \geq \boldsymbol{\kappa}_2$.

Principal normal curvatures: $\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2$ (The maximum and minimum of $Q_{II}$ on the unit circle of $T_{\mathbf{p}}(\mathbf{S})$)

Principal normal directions: $\mathbf{u}_1, \mathbf{u}_2$

Gauss curvature: $K(\mathbf{p})$ = determinant of $D\mathbf{n}(\mathbf{p})$

$$K = \boldsymbol{\kappa}_1\boldsymbol{\kappa}_2 = \frac{LN - M^2}{EG - F^2}$$

Mean curvature: $H(\mathbf{p}) = -\frac{1}{2}\,\mathrm{Trace}(d\mathbf{n}(\mathbf{p}))$

$$H = \frac{\boldsymbol{\kappa}_1 + \boldsymbol{\kappa}_2}{2} = \frac{EN + GL - 2FM}{2(EG - F^2)}$$

Weingarten equations:

$$\mathbf{n}_u = \frac{MF - LG}{EG - F^2}\Phi_u + \frac{LF - ME}{EG - F^2}\Phi_v$$

$$\mathbf{n}_v = \frac{NF - MG}{EG - F^2}\Phi_u + \frac{MF - NE}{EG - F^2}\Phi_v$$

Principal normal curvatures are roots of: $k^2 - 2Hk + K = 0$

$$\boldsymbol{\kappa}_1 = H + \sqrt{H^2 - K}, \quad \boldsymbol{\kappa}_2 = H - \sqrt{H^2 - K}$$

Christoffel symbols: $\Gamma^k_{ij}$

$$\Gamma^1_{11} = \frac{GE_u - 2FF_u + FE_v}{2(EG - F^2)}, \quad \Gamma^1_{12} = \Gamma^1_{21} = \frac{GE_v - FG_u}{2(EG - F^2)}, \quad \Gamma^1_{22} = \frac{2GF_v - GG_u - FG_v}{2(EG - F^2)},$$

$$\Gamma^2_{11} = \frac{2EF_u - EE_v + FE_u}{2(EG - F^2)}, \quad \Gamma^2_{12} = \Gamma^2_{21} = \frac{EG_u - FE_v}{2(EG - F^2)}, \quad \Gamma^2_{22} = \frac{EG_v - 2FF_v + FG_u}{2(EG - F^2)}.$$

Geodesics: Assume that $\eta(t)$ is a regular curve in $\mathbf{S}$ and $\gamma(s)$ is its arc-length parameterization with Frenet frame $(T(s),N(s),B(s))$. If

$$n_{\mathbf{S}}(s) = \mathbf{n}(\gamma(s)) \times T(s),$$

then

$$\gamma''(s) = \boldsymbol{\kappa}(s)N(s) = \boldsymbol{\kappa}_n(s)\mathbf{n}(\gamma(s)) + \boldsymbol{\kappa}_g(s)n_{\mathbf{S}}(s)$$
$$\boldsymbol{\kappa}^2 = \boldsymbol{\kappa}_n^2 + \boldsymbol{\kappa}_g^2$$
$$\boldsymbol{\kappa}_g(s) = \boldsymbol{\kappa}(s)\mathbf{n}(\gamma(s)) \bullet B(s) = \boldsymbol{\kappa}(s)\cos\alpha(s)$$
$$\boldsymbol{\kappa}_g(t) = \frac{\eta''(t) \bullet (\mathbf{n}(\eta(\mathbf{t})) \times \eta'(t))}{|\eta'(t)|^3}.$$

Covariant derivative: $$\frac{D\mathbf{X}}{dt}(t) = \mathbf{X}'(t) - (\mathbf{X}'(t) \bullet \mathbf{n}(\eta(t)))\mathbf{n}(\eta(\mathbf{t}))$$

Exponential map: $\exp_{\mathbf{p}} : \mathbf{U}(\mathbf{p}) \to \mathbf{S}$, where $\exp_{\mathbf{p}}(\mathbf{v}) = \gamma(1)$

Parallel or offset surface: $p_d(u,v) = p(u,v) + \mathbf{dn}(u,v)$

$$(p_d)_v = p_v + d\mathbf{n}_v \quad \text{and} \quad (p_d)_u = p_u + d\mathbf{n}_u$$
$$N_d(u,v) = (1 - 2Hd + Kd^2)N(u,v)$$

Define $\sigma$ by $\mathbf{n}_d = \sigma\,\mathbf{n}$.

$$(\boldsymbol{\kappa}_1)_d = \frac{\sigma\boldsymbol{\kappa}_1}{1+d\boldsymbol{\kappa}_1}, \quad (\boldsymbol{\kappa}_2)_d = \frac{\sigma\boldsymbol{\kappa}_2}{1+d\boldsymbol{\kappa}_2}$$

$$K_d = \frac{K}{1+2Hd+Kd^2}, \quad H_d = \sigma\frac{H+Kd}{1+2Hd+Kd^2}$$

Ruled surface: $p(u,v) = q(u) + v\alpha(u)$

If $\lambda = \dfrac{(q' \times \alpha) \bullet \alpha'}{\alpha' \bullet \alpha'}$, then

$$K = K(u,v) = \frac{-M^2}{EG - F^2} = -\frac{\lambda^2}{(\lambda^2 + v^2)^2}$$

## 9.20 Exercises

### Section 9.2

9.2.1. Find the length of the helix $p(t) = (a\cos t, a\sin t, bt)$, $t \in [0,\pi]$.

9.2.2. Find the arc-length parameterization of the curve $p(t) = (\cosh t, \sinh t, t)$, $t \in [0,3]$.

9.2.3. Find the arc-length parameterization of the curve $p(t) = (e^t\cos t, e^t\sin t, e^t)$, $t \in [0,2]$.

**Section 9.3**

9.3.1. Find the principal normal, the signed curvature, and the center of curvature for the following curves p(t):

(a) (parabola) $p(t) = (t,t^2)$
(a) (cycloid) $p(t) = (t - \sin t, 1 - \cos t)$
(b) (catenary) $p(t) = (t, \cosh t)$

9.3.2. Show that a noncircular ellipse has four vertices.

**Section 9.4**

9.4.1. Prove that the length of a curve, its curvature, and its torsion are invariant under a rigid motion.

9.4.2. Find the curvature, torsion, Frenet frame, and the equations for the osculating, normal, and rectifying plane for the *twisted cubic* $p(t) = (t,t^2,t^3)$ at the origin.

9.4.3. Consider the curve $p(t) = (t - \sin t, 1 - \cos t, t)$. Show that its curvature and torsion is defined by the following formulas:

$$\kappa(t) = \frac{\left(1+4\sin^4\frac{t}{2}\right)^{1/2}}{\left(1+4\sin^2\frac{t}{2}\right)^{3/2}}, \quad \tau(t) = -\frac{1}{1+4\sin^4\frac{t}{2}}.$$

9.4.3. Define a curve in $\mathbf{R}^3$ to be a *generalized helix* if it admits a regular parameterization F(t) with the property that the tangent vector F′(t) makes a constant angle θ with a fixed unit vector **u**, where $0 < \theta < \pi/2$. Prove that a curve is a generalized helix if and only if the ratio $\tau/\kappa$ is a nonzero constant (see [Wein00] or [Lips69]).

9.4.4. Let γ(t) be a curve. Show that if γ′(t) and γ″(t) are linearly dependent for all t, then γ(t) is a straight line.

**Section 9.6**

9.6.1. Show that the equation of the involute of the circle $p(t) = (a\cos t, a\sin t)$, $a > 0$, is

$$q(t) = (a(\cos t + t\sin t), a(\sin t - t\cos t)).$$

9.6.2. Show that the equation of the involute of the catenary $p(t) = (t, \cosh t)$, is

$$q(t) = (t - \sinh t \cosh t, 2\cosh t).$$

**Section 9.7**

9.7.1. Let p(t) be a closed convex plane curve. Let

$$p_d(t) = p(t) + dn(t), \quad d > 0,$$

be a curve parallel to p(t), where n(t) are the unit vectors pointing out of the region enclosed by p(t). Show that

$$\text{Length of } p_d(t) = \text{Length of } p(t) + 2\pi d.$$

**Section 9.8**

9.8.1. Let $\Phi:[a,b] \to \mathbf{C}$ be a regular parameterization of a curve $\mathbf{C}$ in $\mathbf{R}^2$. Show that the volume $V(\Phi)$ of the parameterization $\Phi$ as defined in this section agrees with the definition of the length of a curve given in Section 9.2.

9.8.2. Show by direct computation that Theorem 9.8.5 computes the same length of a curve in the plane as the definition in Section 9.2.

9.8.3. Find the coefficients of the first fundamental form for the surface of revolution that is parameterized by $\Phi(x,\theta) = (f(x)\cos\theta, f(x)\sin\theta, g(x))$.

9.8.4. Consider the parameterization

$$\Psi:\ [0,2\pi] \times (0,\pi) \to \mathbf{R}^2 - \mathbf{0}$$

defined by

$$\Psi(\theta,\phi) = \left(\frac{\cos\theta\,\sin\phi}{1-\cos\phi}, \frac{\sin\theta\,\sin\phi}{1-\cos\phi}\right),$$

which is the composite of the spherical coordinate parameterization of $\mathbf{S}^2$ in Example 9.8.4 and the stereographic projection. Compute the coefficients E, F, and G with respect to $\Psi$. Deduce from this and the computations in Example 9.8.4 that the stereographic projection does not preserve area.

**Section 9.9**

9.9.1. Show that (0,0,0) is a planar point of the surface of revolution where the curve $z = x^4$ is revolved about the z-axis.

9.9.2. Prove that a curve in a surface is an asymptotic line if and only if every point of the curve is an inflection point or the osculating plane at the point is tangent to the surface.

9.9.3. Let $\Phi(u,v)$ be a regular parameterization of a neighborhood of a point $\mathbf{p}$ in a surface $\mathbf{S}$. Prove that the u- and v-parameter curves of the patch, that is, $\Phi(u,0)$ and $\Phi(0,v)$, respectively, are asymptotic lines for the surface if and only if $L = N = 0$ at each point.

9.9.4. Let $\mathbf{S}$ be the surface of revolution obtained by revolving a curve

$$\gamma:\ [0,1] \to \mathbf{R}^2 \quad , \quad \gamma(t) = (x(t), y(t)),$$

about the x-axis. Assume that $y(t) > 0$. Let

$$\varphi:\ [0,1] \times [0,2\pi] \to \mathbf{R}^3 \quad , \quad \varphi(t,\theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta),$$

be the standard parameterization of **S**. Show that

(a) $K = \dfrac{x'(x''y' - x'y'')}{y(x'^2 + y'^2)}$

(b) If $\gamma(t)$ is the arc-length parameterization, then $K = -y''/y$.

(c) Assume that $x(t) = t$. Show that every point of **S** is a parabolic point if and only if **S** is a cylinder ($y(t) = c$) or a cone ($y(t) = at + b$, $a \neq 0$).

9.9.5. Show that a definition of the second fundamental form based on Equation (9.47) is independent of the choice of local coordinates.

9.9.6. Let **p** be a point of a surface **S**. Let L, M, and N be the coefficients of the second fundamental form of **S**. Show that the point **p** is

(1) elliptic if $LN - M^2 > 0$,
(2) hyperbolic if $LN - M^2 < 0$,
(3) parabolic if $LN - M^2 = 0$ but not all of the L, M, and N are zero, and
(4) planar or flat if $L = M = N = 0$.

These characterizations of elliptic, hyperbolic, parabolic, and planar or flat are sometimes used as the definition.

9.9.7. Find the Gauss and mean curvatures of the surface $\Phi(u,v) = (u + v, u - v, uv)$ at (1,1).

### Section 9.10

9.10.1. Prove Theorem 9.10.1(3). See [Stok69].

9.10.2. Show that the curve $\gamma(t) = (t,0,t^2)$ is a geodesic in the surface $\Phi(u,v) = (u,v,u^2 - v^2)$.

9.10.3. Let **S** be a surface in $\mathbf{R}^3$.

**Definition.** Let **X** be a plane in $\mathbf{R}^3$ and let r be the reflection about **X**. We say that **X** is a *plane of symmetry* for **S** if for all points **p** in **S** the reflected point r(**p**) belongs to **S**.

Prove that if **X** is a plane of symmetry for **S**, then the intersection of **X** and **S** defines a geodesic curve. Use this fact to find some geodesics on the surface defined by $x^2 + y^2 - z^2 = 1$.

9.10.4. Prove Theorem 9.10.15.

9.10.5. Prove Theorem 9.10.20.

### Section 9.12

9.12.1. Let $\gamma(s)$ be a space curve parameterized by arc-length. Let (T(s),N(s),B(s)) be its moving trihedron. Consider the canal surface **S** with constant radius function and parameterization

$$\Phi(s,\theta) = \gamma(s) + (r\cos\theta)N(s) + (r\sin\theta)B(s).$$

Let $\kappa$ be the curvature of $\gamma$ and K the Gauss curvature of **S**. Prove that

(a) $g = r^2(1 - \kappa r\cos\theta)^2$

(b) $K = -\dfrac{\kappa\cos\theta}{\sqrt{g}}$

**Section 9.14**

9.14.1. Show that if $\mathbf{S}$ is a surface of constant mean curvature $h \neq 0$, then the parallel surface $\mathbf{S}_{1/2h}$ has constant Gauss curvature $4h^2$. (This is a theorem of Bonnet.)

**Section 9.15**

9.15.1. (a) Prove Lemma 9.15.3.
(b) Consider a curve

$$\mu(u) = q(u) + r(u)\alpha(u).$$

Show that the condition $\mu' \bullet \alpha' = 0$ defines a unique such curve if $\alpha' \neq 0$ and that this is the line of striction.

9.15.2. Find the line of striction for the Moebius strip

$$\mathbf{\Phi}(\theta, t) = \left(\left(2 + t\cos\frac{\theta}{2}\right)\cos\theta, \left(2 + t\sin\frac{\theta}{2}\right)\sin\theta, t\sin\frac{\theta}{2}\right).$$

9.15.3. Show that every singular point of a noncylindrical ruled surface lies on its line of striction.

9.15.4. Show that the surface obtained by revolving a segment about a line is developable.

9.15.5. Let $\gamma(s)$ be a space curve parameterized by arc-length. Let $(T(s),N(s),B(s))$ be its moving trihedron. Find conditions under which the following surfaces are developable:

(a) $\Phi(s,t) = \gamma(s) + tN(s)$
(b) $\Phi(s,t) = \gamma(s) + tB(s)$

9.15.4. Show that the hyperbolic paraboloid $z = x^2 - y^2$ is a doubly ruled surface.

**Section 9.16**

9.16.1. Let $\gamma(s)$ be a curve in $\mathbf{R}^3$ parameterized by arc-length. Assume that $\kappa(s) > 0$. Let $(\mathbf{U}_1,\mathbf{U}_2,\mathbf{U}_3)$ be a frame field that extends the Serret frame $(T,N,B)$ in a neighborhood of the curve. Let $\mathbf{X} = \mathbf{U}_1$. Show that the Equations (9.94) in Theorem 9.16.6 reduce to the Serret-Frenet equations.

CHAPTER 10

# Algebraic Geometry

*Prerequisites:* Chapter 3, 4, Sections 8.3–8.5, 8.11, 8.14, and basic abstract algebra

## 10.1 Introduction

Smooth manifold-like spaces are typically presented in one of two ways: either via an explicit parameterization or as the set of zeros of some function. The definition of a manifold was based on a parameterization, but special cases of manifolds, most notably the conic sections, are described most easily by equations. The functions involved are usually polynomials or at worst rational functions, which are quotients of polynomials. Polynomials define ideals in polynomial rings and so algebraic structures become associated to implicitly defined spaces. The question then arises as to whether there is any connection between purely algebraic properties of substructures of polynomial rings (a subject that belongs to the field of commutative algebra) and geometric properties of associated spaces. This is what algebraic geometry is all about. Algebraic geometry can be thought of as the study of commutative algebra (the "algebraic" part of the name) as seen through the eyes of a geometer (the "geometry" part). It is another attempt to study geometric properties using algebra, just like algebraic topology.

The definitions below introduce the sets and the associated terminology that is basic to algebraic geometry. These definitions and others along with various results in this chapter are naturally expressed in the context of arbitrary fields k and subsets of $k^n$. However, it turns out that the best way to understand what is going on is to analyze what happens in the case of algebraically closed fields first. What sets these fields apart from others is that every polynomial of degree n over such a field has n roots, so that sets defined as zeros of polynomials will always have the "right" number of points. This explains why most theorems in algebraic geometry deal with algebraically closed fields. Drawing the desired conclusions back down at the original field level is an additional and usually nontrivial step. Although we shall phrase much of the discussion in this chapter in terms a general "field k" or "algebraically closed field k," to keep things concrete, the reader can always interpret this to mean either **R** or

its algebraic closure, the complex numbers **C**, respectively. On a practical level we are of course most interested in the field **R** and subsets of $\mathbf{R}^n$.

**Definition.** Let k be any field and let $f(X_1,X_2,\ldots,X_n)$, $f_i(X_1,X_2,\ldots,X_n) \in k[X_1,X_2,\ldots,X_n]$. A set of points in $k^n$ of the form

$$V(f) = \{\mathbf{c} \in k^n \mid f(\mathbf{c}) = 0\}$$

is called a *hypersurface* in $k^n$. If n = 2, then we call V(f) an *algebraic plane curve* (often abbreviated to *plane curve* in this chapter). A set of points in $k^n$ of the form

$$V(\{f_i\}) = \{\mathbf{c} \in k^n \mid f_i(\mathbf{c}) = 0 \text{ for all } i\}$$

is called an *affine algebraic variety*, or simply *variety*, in $k^n$ defined by the set of polynomials $\{f_i\}$. If the field k has to be made explicit, we shall use the terms "k-hypersurface" and "k-variety" and use the notation $V_k(f)$ and $V_k(\{f_i\})$. A *real* or *complex variety* will mean a k-variety where k is **R** or **C**, respectively.

The justification for the term "hypersurface" will be found later in Theorem 10.16.8, which shows that these sets are typically (n – 1)-dimensional subsets of $k^n$.

Algebraic geometry is the study of varieties, but we shall see that thorough answers to many of the questions we want to ask all involve projective space. In fact, the natural setting for most of the topics in this chapter is projective space. Even though we may be motivated by problems about algebraic curves in affine space, the general approach is to study the curves in projective space first and then translate the appropriate answers back to the affine context. *Affine space*, namely, $k^n$, should be thought of as dealing with local properties of curves, whereas projective space deals with their intrinsic properties that are both local **and** global. We need to extend the definitions above to that setting. In fact, we need to generalize our earlier definition of projective space (as given in Section 3.4).

**Definition.** Define *n-dimensional projective space* $\mathbf{P}^n(k)$ *over a field k* to be the quotient space

$$\mathbf{P}^n(k) = (k^{n+1} - \mathbf{0})/\sim,$$

where ~ is the equivalence relation on $k^{n+1} - \mathbf{0}$ defined by

$$(c_1,c_2,\ldots,c_{n+1}) \sim (dc_1,dc_2,\ldots,dc_{n+1})$$

for all $c_i$, $d \in k$, and $d \neq 0$. The equivalence class of an n-tuple $\mathbf{c} = (c_1,c_2,\ldots,c_{n+1})$ will be denoted by $[\mathbf{c}]$ or $[c_1,c_2,\ldots,c_{n+1}]$.

Note that the space $\mathbf{P}^n$, as defined in Section 3.4, is just $\mathbf{P}^n(\mathbf{R})$, the important special case of n-dimensional projective space over the reals.

To define a concept of an algebraic variety in $\mathbf{P}^n(k)$, meaning the zeros of polynomials, one has to be careful. The problem is that if $f(X_1,X_2,\ldots,X_{n+1})$ is a polynomial in $k[X_1,X_2,\ldots,X_{n+1}]$, then the obvious definition of the value of f at a point **p** in $\mathbf{P}^n(k)$, namely,

$$f(\mathbf{p}) = f(c_1, c_2, \ldots, c_{n+1}) \quad \text{if} \quad \mathbf{p} = [c_1 c_2, \ldots, c_{n+1}],$$

is not well defined in general because it depends on the representative tuple chosen for **p**. On the other hand, that definition **is** well defined (and we shall use it) with respect to the statements

$$f(\mathbf{p}) = 0 \quad \text{or} \quad f(\mathbf{p}) \neq 0$$

**provided** that we restrict ourselves to homogeneous polynomials. This follows from the fact that if f is a homogeneous polynomial of degree d, then

$$f(t\mathbf{c}) = t^d f(\mathbf{c}),$$

so that if f is zero or nonzero on one representative **c** of **p**, then it will be zero or nonzero on all of them.

**Definition.** Let $f(X_1, X_2, \ldots, X_{n+1})$ be a homogeneous polynomial in $k[X_1, X_2, \ldots, X_{n+1}]$. Define the *hypersurface* V(f) in $\mathbf{P}^n(k)$ by

$$V(f) = \{\mathbf{p} \in \mathbf{P}^n(k) \mid f(\mathbf{p}) = 0\}.$$

If n = 2, then we shall call V(f) a *projective algebraic plane curve* (again often abbreviated to *plane curve*). If f is a linear homogeneous polynomial, then V(f) will be called a *hyperplane*. If n = 2, then such a set will also be called a *line*. Given a finite collection of homogeneous polynomials $f_i(X_1, X_2, \ldots, X_{n+1})$ in $k[X_1, X_2, \ldots, X_{n+1}]$, the *(projective) algebraic variety*, or simply *variety*, in $\mathbf{P}^n(k)$ defined by the set of polynomials $f_i$ is the set

$$V(\{f_i\}) = \{\mathbf{p} \in \mathbf{P}^n(k) \mid f_i(\mathbf{p}) = 0 \text{ for all } i\}.$$

Again, if the field k has to be made explicit, we shall refer to "k-hypersurfaces" and "k-varieties" and use the notation $V_k(f)$ and $V_k(\{f_i\})$.

(We note that another common approach to projective varieties is described later in Exercise 10.8.5.)

This chapter attempts to give a brief overview of some central results about the structure of varieties. As with some of the other topics covered in previous chapters, we shall only barely scratch the surface of the field of algebraic geometry. Our emphasis will be on plane curves, but even for this special case much will be left out. Therefore, the main goal will be to motivate and organize various important topics for the reader. By and large proofs are only sketched and many are omitted, but at least he/she will at least see some of what is known. The problem is that to do more would require a much more in-depth knowledge of commutative algebra than we assume here. The field of algebraic geometry has much more advanced mathematics as a prerequisite than any of the other topics discussed in this book. It is hoped, however, that the reader will get a feel for the subject and will be motivated to refer to some of the references to learn more. Certainly, by the time the chapter ends, he/she should have a better understanding of the following problems that are of particular relevance to CAGD:

(1) (The Implicitization Problem) Given a polynomial parameterization p(t) of a space **X**, can one find an implicit equation which defines **X**?
(2) (The Parameterization Problem) If a space **X** is defined implicitly by means of a polynomial equation f(**p**) = 0, can one parameterize **X** using polynomials?
(3) (The Intersection Problem) Given implicit or parametric definitions for spaces **X** and **Y**, what can be said about their intersection $\mathbf{X} \cap \mathbf{Y}$?

Here is a look at what is to come in this chapter. After Section 10.2 gets us started with some examples of plane curves that show why affine space is inadequate for a thorough analysis of varieties, Sections 10.3–10.5 back up to fill in some needed algebraic background. Section 10.3 describes some analytic properties of the parameterization of projective space and useful facts about how to pass back and forth between affine and projective space. Section 10.4 defines the resultant of two polynomials and shows how it can be used to find common factors. Section 10.5 describes some basic algebraic properties of polynomials and their influence on the structure of varieties. Sections 10.6 and 10.7 define intersection multiplicities, singularities, and tangents of plane curves and use this to analyze their intersections. Next, as preparation for studying higher-dimensional varieties, we develop some simple aspects of commutative algebra in Section 10.8. Section 10.9 looks at the problem of finding implicit representations of parametrically presented curves. Difficulties with using the resultant lead to a discussion of Gröbner bases in Section 10.10 and elimination theory in Section 10.11. Section 10.12 starts the analysis of the singularities of a curve and defines the place of a curve. This is a long section because in order to understand what is going on we have to bring complex analysis into the picture, in particular, the topics of analytic continuation, the uniformization problem, and Riemann surfaces. Section 10.13 is on rational and birational maps. We move on to higher dimensions and space curves in Section 10.14. The parameterization problem for implicitly defined curves is discussed in Section 10.15. We finish the chapter with an overview of some higher-dimensional topics in Sections 10.16–10.18. We define the dimension of a variety, describe the Grassmann manifolds as varieties, and sketch some important theorems.

**A final note.** At various times in this chapter we will be taking derivatives or partial derivatives of polynomials. Let us clarify this right now, so that the reader will not be puzzled by what that might mean in the case of, say, polynomials over the complex numbers or other fields. In the case of polynomials there is a formal notion of derivative that is computed like the usual derivative but is well defined for polynomials over any ring and does not involve having to take limits. See Section B.7. This is what we shall be using.

## 10.2 Plane Curves: There Is More than Meets the Eye

We begin our tour of algebraic geometry with a closer look at plane curves. Even though the reader's initial reaction might be that their low dimensionality would not lead to anything interesting, this is not at all the case. In fact, in contrast to algebraic topology where things tend to get interesting only in higher dimensions, an analysis of plane curves will already lead to some of the most fundamental ideas in algebraic

geometry. The only thing simple about the two-dimensional case is the fact that it is somewhat easier to draw pictures here and so the reader will have an easier time understanding what some of the relevant issues are. This section will show the potential richness of a theory of plane curves by attempting to answer the simple sounding question: what do plane curves and their intersections "really" look like?

The most well-known plane curves are probably the conics, namely, the ellipse, parabola, and hyperbola, which are defined by simple quadratic equations. As we look at such curves let us do it in the context of the following question:

> What is the relation between the dimension of a variety and the number and type of equations that define it?

A natural intuition would be that each equation takes away one degree of freedom. Certainly, in the case of linear equations, k "independent" linear equations in n variables will define an (n – k)-dimensional plane in $\mathbf{R}^n$, or, to put it another way, a plane of codimension k. This uniformity does not seem to be the case when dealing with polynomial equations. Consider, for example, the three polynomials:

$$X^2 + Y^2 - 1 \tag{10.1}$$

$$X^2 + Y^2 \tag{10.2}$$

$$X^2 + Y^2 + 1 \tag{10.3}$$

The curves in $\mathbf{R}^2$ associated to the zeros of these polynomials are the circle, the point $\mathbf{0}$, and the empty set, respectively. Although the circle is a codimension 1 subset, the other two spaces are not. One might be tempted to give up on the general principle that one equation should define a codimension 1 subset, but that would be premature. One can restore a uniformity into this picture by doing what is often done in mathematics – extending the universe in which one is searching for solutions. In this case, the problem is that the reals are not algebraically closed. We should consider our polynomials as polynomials over the complex numbers $\mathbf{C}$. This will guarantee that there will always be the "right" number of solutions. In fact, one should go one step further and move to homogeneous coordinates and complex projective space $\mathbf{P}^2(\mathbf{C})$. This will restore some additional "missing" points "at infinity" and unify connectivity properties of varieties. For example, we showed in Chapter 3 that all the conics look essentially the same and the differences between ellipses, parabolas, and hyperbolas that exist in the affine world when it comes to their number of branches is eliminated. So let us show how using complex numbers and projective space does bring back harmony and uniformity to our world. Of course, the reader should note that the term "plane curve" has now taken on a new meaning. The curve, as the set of zeros of a polynomial in two variables whose coefficients might be all real, can mean the "real part" in $\mathbf{R}^2$ **or** the "complex part" in $\mathbf{C}^2$ **or** the entire curve in $\mathbf{P}^2(\mathbf{C})$. We shall switch back and forth between these meanings. The context should make clear which we have in mind at any given time.

What happens when X and Y are complex variables rather than real ones? We look at several examples starting with the polynomials in equations (10.1–3). See [Kend77] for more details about the examples below.

**10.2.1. Example.** V $(X^2 + Y^2 - 1)$ (The complex "circle")

**Analysis.** Replacing X by u + **i**v and Y by z + **i**w gives the equation

$$(u+\mathbf{i}v)^2+(z+\mathbf{i}w)^2=1.$$

Setting the real and complex parts of both sides of the equation equal gives

$$u^2-v^2+z^2-w^2=1. \tag{10.4}$$

and

$$uv+zw=0. \tag{10.5}$$

We shall try to analyze the space $V(X^2 + Y^2 - 1)$ by thinking of it as the set of (u,v,z,w)-tuples in $\mathbf{R}^4$ defined by equations (10.4) and (10.5) and look at various sections of it. First, let us look at the three-dimensional slice w = 0. We get two equations

$$u^2-v^2+z^2=1 \tag{10.6}$$

and

$$uv=0. \tag{10.7}$$

Let $\mathbf{C}_a(e)$ denote the circle in the plane v = a defined by

$$\mathbf{C}_a(e)=\{(u,a,z)\mid u^2+z^2=1+a^2+e^2\},$$

and let $\mathbf{C}_a = \mathbf{C}_a(0)$. Let **H** be the hyperbola in the z-v plane defined by

$$z^2-v^2=1.$$

Then it is easy to see that the surface defined by equation (10.6) is just the union of the circles $\mathbf{C}_a$ or, alternatively, the surface obtained by revolving the hyperbola **H** about the v axis. See Figure 10.1(a).

Equation (10.7) defines two planes (u = 0 and v = 0). The intersection in $\mathbf{P}^2(\mathbf{C})$ of the hyperboloid defined by (10.6) and these two planes is shown in Figure 10.1(b), where $\mathbf{p}_\infty = [1,-1,0]$ and $\mathbf{p}'_\infty = [1,1,0]$ denote the two points "at infinity" associated to **H**. Geometrically, the points $\mathbf{p}_\infty$ and $\mathbf{p}'_\infty$ come from the two asymptotes of the hyperbola defined by the lines x = y and x = –y.

Next, consider the three-dimensional slice w = e. The two equations we have to look at now are

$$u^2-v^2+z^2=1+e^2 \tag{10.8}$$

and

$$z=\left(-\frac{v}{e}\right)u. \tag{10.9}$$

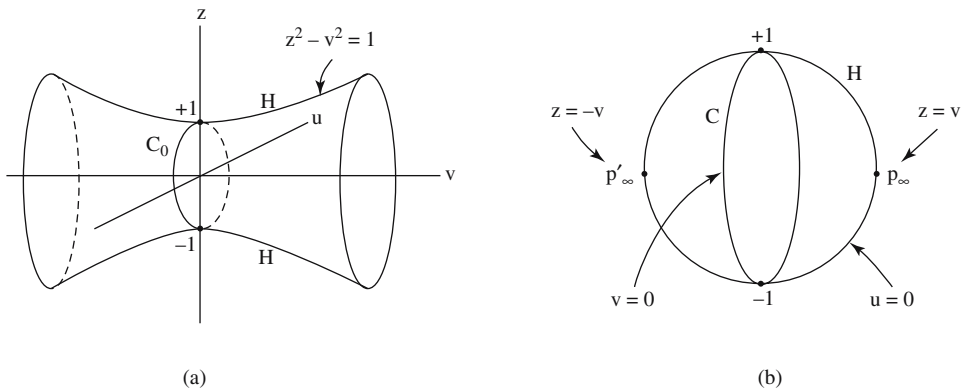



**Figure 10.1.** Example 10.2.1.

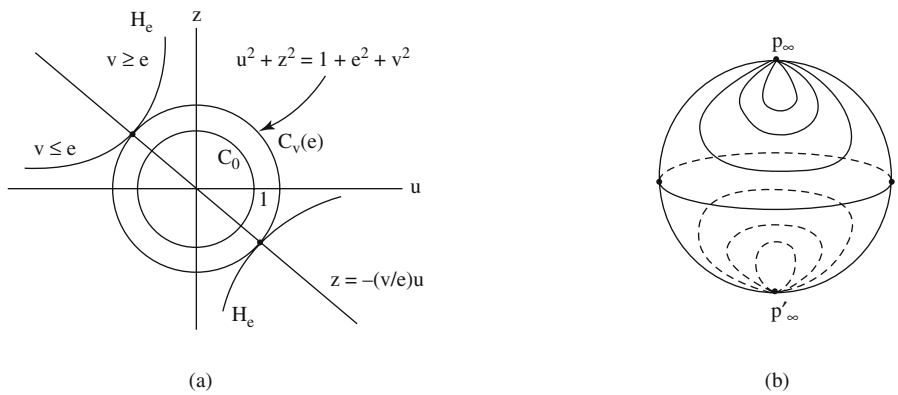



**Figure 10.2.** Example 10.2.1 continued.

Equation (10.8) by itself defines hyperboloids $\mathbf{H}_e$, which are the union of the circles $\mathbf{C}_v(e)$, of which $\mathbf{H}$ in Figure 10.1(a) is a special case. Equation (10.9) defines a twisted plane. For fixed v it defines a line. These lines start with slope 0 and then rotate to almost vertical lines as v gets much bigger than 1. Figure 10.2(a) shows the circles $\mathbf{C}_v(e)$ and hyperbolas $\mathbf{H}_e$ defined by equations (10.8) and (10.9) projected to the u-z plane. One can show that the curves defined by the intersection of the hyperboloids and twisted planes (the simultaneous solutions to equations (10.8) and (10.9)) "fill out" the skeleton of the sphere shown in Figure 10.1(b). See Figure 10.2(b). In other words, thought of as living in $\mathbf{P}^2(\mathbf{C})$ the space $V(X^2 + Y^2 - 1)$ is topologically a sphere.

**10.2.2. Example.** $V(X^2 + Y^2)$

**Analysis.** Again replacing X and Y by u + **i**v and z + **i**w, respectively, leads to the equations

$$u^2 - v^2 + z^2 - w^2 = 0 \tag{10.10}$$

and

$$uv + zw = 0. \tag{10.11}$$

The three-dimensional slice w = 0 is defined by

$$u^2 - v^2 + z^2 = 0 \tag{10.12}$$

and

$$uv = 0. \tag{10.13}$$

Equation (10.12) defines a cone. Equation (10.13) defines two planes as before. See Figure 10.3(a). What basically is happening is that if we were to consider the family of hypersurfaces $V(X^2 + Y^2 - c)$, $c \in [0,1]$, and let c approach 0, then we would get pictures like in Figure 10.1 except that the circle $\mathbf{C}_0$ would shrink to $\mathbf{0}$ and the lines $\mathbf{H}$ would "straighten" out. The final analog of Figure 10.2(b) would be Figure 10.3(b). In other words, thought of as living in $\mathbf{P}^2(\mathbf{C})$ the space $V(X^2 + Y^2)$ is topologically the union of two spheres which meet at a point.

**10.2.3. Example.** $V(X^2 + Y^2 + c)$, $0 \neq c \in \mathbf{C}$

**Analysis.** One can show that the closure of this space in $\mathbf{P}^2(\mathbf{C})$ is topologically a sphere (Exercise 10.2.1).

**10.2.4. Example.** $V(Y^2 - X(X^2 - 1))$

**Analysis.** This time, replacing X by u + **i**v and Y by z + **i**w leads to the equations

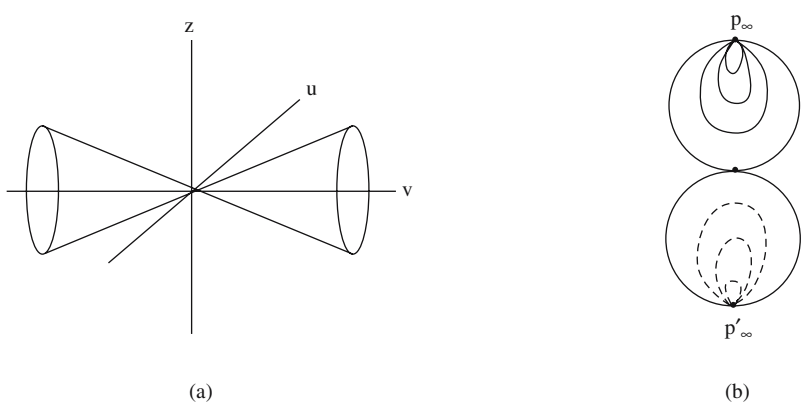



**Figure 10.3.** Example 10.2.2.

$$z^2 - w^2 = u^3 - 3uv^2 - u \tag{10.14}$$

and

$$2zw = 3u^2v - v^3 - v. \tag{10.15}$$

When v = 0, this reduces to

$$z^2 - w^2 = u^3 - u \tag{10.16}$$

and

$$zw = 0. \tag{10.17}$$

If w = 0, then we get the equation

$$z^2 = u^3 - u.$$

The graph of this equation is shown in Figure 10.4(a). If z is set to zero instead of w, then we get a mirror image of this graph and so the complete picture for the case v = 0, as specified by equations (10.16) and (10.17), is as shown in Figure 10.4(b). Because the tangent lines of the two branches approach the z and w axis, respectively, and since the z-w plane corresponds to a unique point $\mathbf{p}_\infty$ in $\mathbf{P}^2(\mathbf{C})$, the graphs look topologically as shown in Figure 10.5(a). By next considering the three-dimensional slices v = e, one can show that they fill a torus. See Figure 10.5(b). Therefore, thought of as living in $\mathbf{P}^2(\mathbf{C})$ the space $V(Y^2 - X(X^2 - 1))$ is topologically a torus.

After these examples of projective complex plane curves, is there an underlying general result? Yes!

**10.2.5. Theorem.** Let f(X,Y) be a nonconstant polynomial in $\mathbf{C}[X,Y]$. If f is irreducible, then the closure of V(f) in $\mathbf{P}^2(\mathbf{C})$ is a compact connected orientable surface

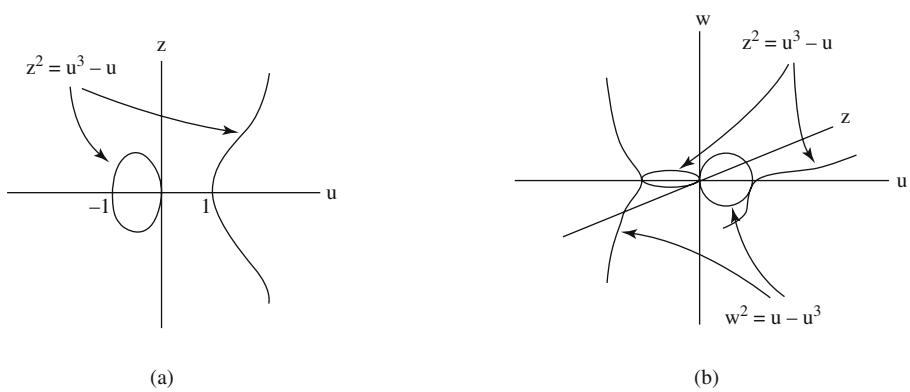



**Figure 10.4.** Example 10.2.4.

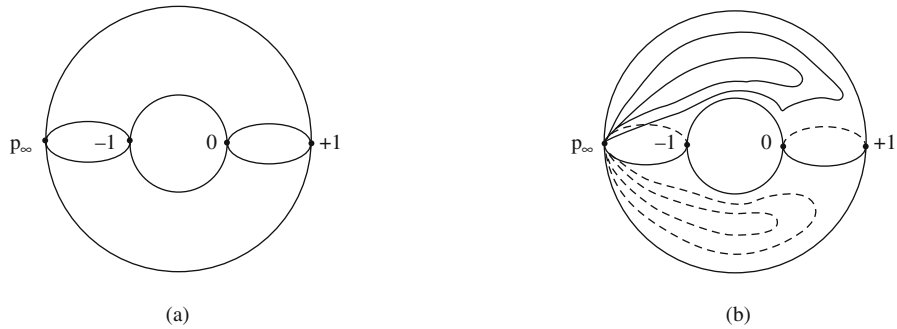



**Figure 10.5.** Example 10.2.4.

in which possibly a finite number of pairs of points have been identified. In general, the closure of V(f) is a finite union of such spaces where each is attached to every other along a finite set of points.

**Proof.** See [Kend77].

Theorem 10.2.5 tells us what an arbitrary algebraic curve in $\mathbf{P}^2(\mathbf{C})$ looks like topologically, but it does **not** say that every surface of the type described in the theorem is an algebraic curve, only some are. However, one can use the polynomial

$$Y^2 - X(X^2 - 1)\ldots(X^2 - g^2)$$

to generate a space of genus g with 2-to-1 identifications. Constructing spaces with more complicated identifications gets more complicated.

We have just seen that if we want to understand plane curves there is more than what meets the eye at first glance. A similar revelation can be had with respect to the problem of plane curve intersections. Two distinct lines (degree 1 curves) can intersect in a single point or none at all. A line and a conic can intersect in 2, 1, or 0 points. Figure 10.6 shows some possible intersections of two distinct conics (degree 2 curves). One can have anywhere from 0 to 4 points in the intersection. At first glance it would appear that there is no connection between the degree of the curves and the number of points in their intersection. However, there **is** a connection. The problem again is that the reals are not algebraically closed. If we think of the curves as lying in $\mathbf{P}^2(\mathbf{C})$, Bézout's theorem (Theorem 10.7.4) will tell us that there are as many points in the intersection, if counted with their multiplicities, as the product of the degrees of the curves. For example, in $\mathbf{P}^2$ all distinct lines intersect in a point. Also, in $\mathbf{P}^2$, the line that corresponds to the y-axis in $\mathbf{R}^2$ and the conic that corresponds to the parabola y = x intersect at the origin and the ideal point [0,1,0].

These preliminary observations should be an indication of the rich theory awaiting us and we now proceed to a more rigorous analysis. Hopefully, the examples in this section have justified our earlier comments about the need to understand things at the complex number and projective space level. Life is easier there and there is

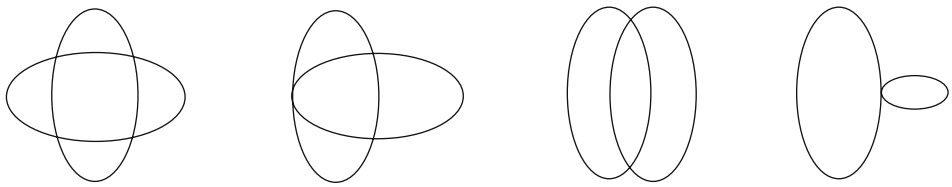


**Figure 10.6.** How conics can intersect.

more uniformity. Of course, in practice one probably wants answers with real numbers, but we shall see over and over again in this chapter that, in so far as it is possible to give such answers, they are obtained by first dealing with the problems over the complex numbers. Before we get down to business though, we need to clarify the connection between affine and projective algebraic geometry and also cover some algebraic preliminaries.

## 10.3 More on Projective Space

This section discusses some important aspects of projective space that were not addressed before because they were not needed until now. We also show how to pass back and forth between projective and affine space.

When $k = \mathbf{R}$ or $\mathbf{C}$, then the space $\mathbf{P}^n(k)$ is actually a differentiable manifold of dimension n or 2n depending on whether k is $\mathbf{R}$ or $\mathbf{C}$, respectively. In fact, $\mathbf{P}^n(\mathbf{C})$ is what is called an n-dimensional **complex manifold**, but since its complex manifold structure will not play any role in this book, we will have nothing further to say about complex manifolds as such. It is important to note though that $\mathbf{P}^n(k)$ looks locally just like $k^n$. More precisely, we can use equations similar to equations (8.45) and (8.46) in Section 8.13 to show that the space can be covered by coordinate neighborhoods $(\mathbf{U}_i,\varphi_i)$, where, for $i = 1, 2, \ldots, n + 1$,

$$\mathbf{U}_i = \{[c_1, c_2, \ldots, c_{n+1}] \mid c_i \neq 0\} \tag{10.18}$$

and

$$\varphi_i\colon\ \mathbf{U}_i \to k^n$$

is defined by

$$\varphi_i([c_1, c_2, \ldots, c_{n+1}]) = \left(\frac{c_1}{c_i}, \ldots, \frac{c_{i-1}}{c_i}, \frac{c_{i+1}}{c_i}, \ldots, \frac{c_{n+1}}{c_i}\right). \tag{10.19}$$

By identifying $k^n$ with $\mathbf{U}_i$ via the homeomorphism $\varphi_i$, one can think of $\mathbf{U}_i$ as consisting of the "affine" points of $\mathbf{P}^n(k)$ with respect to the coordinate neighborhood $(\mathbf{U}_i,\varphi_i)$. The set $\mathbf{P}^n(k) - \mathbf{U}_i$, which is the hyperplane V(f), where

$$f(X_1, X_2, \ldots, X_{n+1}) = X_i,$$

is then referred to as the "plane at infinity" with respect to $(\mathbf{U}_i, \varphi_i)$. The coordinate neighborhood $(\mathbf{U}_{n+1}, \varphi_{n+1})$ represents the "standard" imbedding of $k^n$ in $\mathbf{P}^n(k)$. Although the coordinate neighborhoods $(\mathbf{U}_i, \varphi_i)$ are quite adequate for most situations, they are not for algebraic geometry. An arbitrary subvariety of $\mathbf{P}^n(k)$ might have lots of interesting points lying on its plane at infinity. It would be much more convenient to be able to choose a parameterization with respect to which all of those points, or at least all but a finite number, are affine points. In addition, the choice of a coordinate system clearly has a big influence on the equation for a set, so that the more choices one has, the better off one is. For example, the standard coordinate axes of the projective plane are defined by

$$XY = 0,$$

but they could also be defined by

$$X^2 - Y^2 = 0$$

in another coordinate system. Therefore, we need to rethink our representation of projective space.

From now on, let us think of $\mathbf{P}^n(k)$ as an abstract topological space with its definition inducing only one out of many possible homogeneous coordinates for its points. In fact, let $A = (a_{ij})$ be a nonsingular $(n + 1) \times (n + 1)$ matrix over k and let

$$T: k^{n+1} \to k^{n+1}$$

be the associated linear transformation that sends $\mathbf{x}$ to $\mathbf{x}A$. The map T induces a well-defined map

$$[T]: \mathbf{P}^n(k) \to \mathbf{P}^n(k)$$

defined by $[T]([\mathbf{c}]) = [T(\mathbf{c})]$. Let $T_i$ be the ith component function of T defined by

$$T(\mathbf{c}) = (T_1(\mathbf{c}), T_2(\mathbf{c}), \ldots, T_{n+1}(\mathbf{c})).$$

Then

$$\mathbf{H}_{A,i} = \mathbf{H}_{T,i} = \ker(T_i) \tag{10.20}$$

is a hyperplane in $k^{n+1}$. Finally, let

$$\mathbf{U}_{A,i} = \mathbf{U}_{T,i} = \mathbf{P}^n(k) - \mathbf{H}_{A,i} \tag{10.21}$$

and define

$$\varphi_{A,i} = \varphi_{T,i}: \mathbf{U}_{A,i} \to k^n$$

by

$$\varphi_{A,i}([\mathbf{c}]) = \left(\frac{T_1(\mathbf{c})}{T_i(\mathbf{c})}, \ldots, \frac{T_{i-1}(\mathbf{c})}{T_i(\mathbf{c})}, \frac{T_{i+1}(\mathbf{c})}{T_i(\mathbf{c})}, \ldots, \frac{T_{n+1}(\mathbf{c})}{T_i(\mathbf{c})}\right). \tag{10.22}$$

**10.3.1. Proposition.** The maps $\varphi_{A,i}$ are homeomorphisms and the $(\mathbf{U}_{A,i}, \varphi_{A,i})$ are coordinate neighborhoods that cover $\mathbf{P}^n(k)$.

**Proof.** Exercise 10.3.1.

**Definition.** The tuple $T(\mathbf{c})$ will be called *homogeneous coordinates* of $[\mathbf{c}]$ in $\mathbf{P}^n(k)$ *relative to A*. The collection of coordinate neighborhoods

$$\{(\mathbf{U}_{A,i}, \varphi_{A,i})\}$$

is called the *coordinate neighborhood cover induced by A (or T)*. A point of $\mathbf{H}_{A,i}$ is called an *ideal point* or *point at infinity* of $\mathbf{P}^n(k)$ *with respect to the coordinate neighborhood* $(\mathbf{U}_{A,i}, \varphi_{A,i})$. The set $\mathbf{H}_{A,i}$ is called the *plane* (or *line*, if $n = 2$) *at infinity with respect to the coordinate neighborhood* $(\mathbf{U}_{A,i}, \varphi_{A,i})$.

If A is the identity matrix I, then we get our standard homogeneous coordinates, neighborhoods, ideal points, and planes at infinity for $\mathbf{P}^n(k)$. It will be convenient to abbreviate the above notation as follows:

$$\begin{aligned} \mathbf{H}_i &= \mathbf{H}_{I,i}, \\ \mathbf{U}_i &= \mathbf{U}_{I,i}, \quad \text{and} \\ \varphi_i &= \varphi_{I,i}. \end{aligned} \tag{10.23}$$

In fact, a consequence of the next two theorems is that we may always assume, without loss of generality, that the homogeneous coordinates we use correspond to the most standard of all cases, namely, $i = n + 1$.

**10.3.2. Theorem.** Given any hyperplane in $\mathbf{P}^n(k)$ we can always find a nonsingular $(n + 1) \times (n + 1)$ matrix A so that this hyperplane is the plane at infinity with respect to the coordinate neighborhood $(\mathbf{U}_{A,n+1}, \varphi_{A,n+1})$.

**Proof.** The theorem is an immediate consequence of the fact that given any hyperplane $\mathbf{X}$ in $k^{n+1}$ we can find a nonsingular linear transformation of $k^{n+1}$ that maps $\mathbf{X}$ onto the hyperplane $x_{n+1} = 0$.

When we defined projective hypersurfaces and varieties earlier, those definitions implicitly assumed the **standard** parameterization of projective space. Although a variety is a unique subset of projective space, how it is presented depends on the parameterization. The next theorem tells us how the polynomials that define a variety change as we switch from one set of homogeneous coordinates to another.

**10.3.3. Theorem.** Let $\mathbf{V} = V(f)$ be any hypersurface in $\mathbf{P}^n(k)$ defined by a homogeneous polynomial f. If A is any nonsingular $(n + 1) \times (n + 1)$ matrix, then $\mathbf{V} = V(g)$, where

$$g(Y_1, Y_2, \ldots, Y_{n+1}) = f\big((Y_1, Y_2, \ldots, Y_{n+1})A^{-1}\big),$$

when expressed in homogeneous coordinates relative to A.

**Proof.** The proof is easy. (The reader may also find it helpful to review the discussion in Section 6.10 in [AgoM05] about the way that equations for implicitly defined objects transform.)

**10.3.4. Example.** To find a coordinate neighborhood $(\mathbf{U},\varphi)$ for $\mathbf{P}^2$, so that the line defined by the equation $-2X + Y - Z = 0$ becomes its line at infinity.

**Solution.** Consider the linear transformation

$$\begin{aligned} T:\ X' &= \qquad\quad -Z \\ Y' &= \qquad Y - Z \\ Z' &= -2X + Y - Z \end{aligned}$$

It is nonsingular because the inverse of T is easily found by back substitution to be

$$\begin{aligned} T^{-1}:\ X' &= \quad \frac{1}{2}Y - \frac{1}{2}Z \\ Y' &= -X + Y \\ Z' &= -X \end{aligned}$$

Note that T maps the points (X,Y,Z) with $-2X + Y - Z = 0$ onto the set of points $(X',Y',Z')$ with $Z' = 0$. Therefore, we can let $\mathbf{U} = \mathbf{U}_{T,3}$ and $\varphi = \varphi_{T,3} : \mathbf{U} \to \mathbf{R}^2$, that is,

$$\varphi([X,Y,Z]) = \left(-\frac{Z}{-2X+Y-Z}, \frac{Y-Z}{-2X+Y-Z}\right).$$

The important consequence of Theorem 10.3.3 is that if two coordinate systems for projective space differ by a linear change of variables, then the polynomials that define a variety with respect to these coordinate systems only differ by a linear change of variables. Since a linear change of variables will never affect any of the properties of varieties we are interested in, it follows that by choosing an appropriate homogeneous coordinate system we can reduce problems about properties of points on hypersurfaces in $\mathbf{P}^n(k)$ to problems about points on hypersurfaces in $k^n$. More generally, we shall feel free to choose appropriate coordinate systems for $k^n$ because there is a natural correspondence between linear changes of coordinate systems there and in $\mathbf{P}^n(k)$. These comments justify future phrases like "Without loss of generality assume a coordinate system so that. . . ."

We have indicated that projective space is the natural space in which to do algebraic geometry. However, many problems arise from an attempt to understand affine varieties. This leads to another problem. How do we convert an affine variety into a projective one? Before we address this question we need some more definitions.

Consider a polynomial $f(X_1,X_2, \ldots ,X_n)$ in $k[X_1,X_2, \ldots ,X_n]$. We express f in terms of its homogeneous components. If f has degree d, then f can be written uniquely in the form

$$f = f_0 + f_1 + \ldots + f_d,$$

where $f_i$ is a homogeneous polynomial of degree i in $k[X_1, X_2, \ldots, X_n]$.

**Definition.** The polynomial

$$F(X_1, X_2, \ldots, X_{n+1}) = f_0 X_{n+1}^d + f_1 X_{n+1}^{d-1} + \ldots + f_{d-1} X_{n+1} + f_d$$

in $k[X_1, X_2, \ldots, X_{n+1}]$ is homogeneous of degree d. It is called the *homogenization of f* and is denoted by H(f).

**10.3.5. Example.** If

$$f(X, Y) = 2X^3 - 5XY + XY^2 - 17,$$

then the homogenization of f is

$$F(X, Y, Z) = 2X^3 - 5XYZ + XY^2 - 17Z^3.$$

Another way of looking at this process is that F(X,Y,Z) is obtained by replacing X and Y in f(X,Y) by X/Z and Y/Z, respectively, and then clearing denominators.

**Definition.** If $F(X_1, X_2, \ldots, X_{n+1})$ is a homogeneous polynomial in $k[X_1, X_2, \ldots, X_{n+1}]$, then the polynomial

$$f(X_1, X_2, \ldots, X_n) = F(X_1, X_2, \ldots, X_n, 1) \in k[X_1, X_2, \ldots, X_n]$$

is called the *dehomogenization* of F and is denoted by D(F).

Taking another look at Example 10.3.5 we see that f(X,Y) = F(X,Y,1). In general, D(H(f)) = f but H(D(F)) does **not** always equal F as is shown by the example $F = X_1X_2$ and n = 2.

**Definition.** Let **V** be an (affine) variety in $k^n$. Using the notation defined by equations (10.23), the smallest projective variety in $\mathbf{P}^n(k)$ which contains $\varphi_{n+1}^{-1}(\mathbf{V})$ is called the *projective completion* of **V** relative the coordinate system $(\mathbf{U}_{n+1}, \varphi_{n+1})$ and is denoted by H(**V**). More precisely,

H(**V**) = the intersection of all projective varieties in $\mathbf{P}^n(k)$ that contain $\varphi_{n+1}^{-1}(\mathbf{V})$.

If **V** is a (projective) variety in $\mathbf{P}^n(k)$, then $\mathbf{V} \cap \mathbf{U}_{n+1}$ is called the *affine part* of **V** and will be denoted by D(**V**).

It is the operations of homogenization, dehomogenization, projective completion, and affine part that allow us to pass back and forth between varieties in affine and projective space. See [Kend77] for more details on the relationship between affine and projective varieties. In particular.

**10.3.6. Theorem.** Let $f \in \mathbf{C}[X_1, X_2, \ldots, X_n]$ and $\mathbf{V} = V(f) \subseteq \mathbf{C}^n$.

(1) The (projective) hypersurface $V(H(f))$ in $\mathbf{P}^n(\mathbf{C})$ is the projective completion $H(\mathbf{V})$ of $\mathbf{V}$.
(2) The projective completion $H(\mathbf{V})$ of $\mathbf{V}$ is the topological closure of $\mathbf{V}$ in $\mathbf{P}^n(\mathbf{C})$.
(3) Using the notation defined by equations (10.23), $\varphi_{n+1}^{-1}(\mathbf{V}) = D(H(\mathbf{V}))$. (Basically, this says that $\mathbf{V}$ is the affine part of the projective completion of $\mathbf{V}$.)

**Proof.** See [Kend77] or [Shaf94].

The next example shows that Theorem 10.3.6 is **false** if the field is the reals. The algebraic closure property of the complex numbers is essential.

**10.3.7. Example.** Consider

$$f(X,Y) = Y^2 - X^2(X-1)$$

and let $\mathbf{V} = V(f) \subseteq \mathbf{R}^2$. Figure 10.7(a) shows $\mathbf{V}$. Note that the origin is an isolated point of the graph. It is basically such an isolated point that will lead to our counterexample but it will not be $\mathbf{V}$ directly because we need a variety that has its isolated point at infinity. To get this variety we simply move $\mathbf{V}$. Consider the transformation

$$\begin{aligned} T:\ X' &= Z \\ Y' &= Y \\ Z' &= X, \end{aligned}$$

which moves the y-axis to the line at infinity. This will transform

$$H(f) = Y^2Z - X^3 + X^2Z$$

into

$$G(X,Y,Z) = Y^2X - Z^3 + Z^2X$$

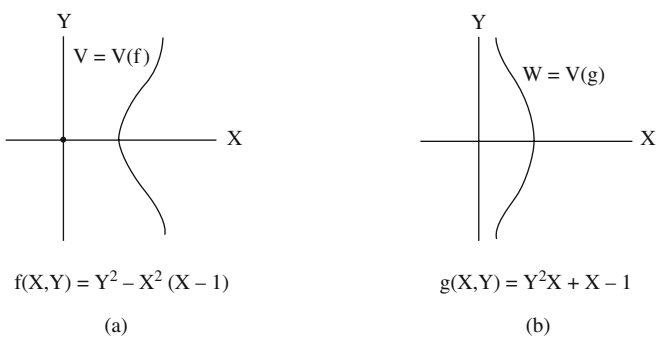



**Figure 10.7.** The varieties of Example 10.3.7.

with

$$g(X,Y) = D(G) = Y^2X + X - 1.$$

Let $\mathbf{W} = V(g) \subseteq \mathbf{R}^2$. Figure 10.7(b) shows $\mathbf{W}$. The variety $\mathbf{W}$ is the counterexample we are looking for. Note that $G = H(g)$ and $[1,0,0] \in V(H(g))$, but the topological closure of $\mathbf{W}$ in $\mathbf{P}^2$ is $\mathbf{W} \cup \{[0,1,0]\}$ (Exercise 10.3.4) and $[1,0,0]$ is not in it. Although we do not yet have the tools to prove this, it is the case that $H(\mathbf{W}) = V(H(g))$ (Exercise 10.5.3). We have found a counterexample to Theorem 10.3.6(2). Actually, $\mathbf{W}$ serves as our counterexample no matter what might have happened because if it were the case that $H(\mathbf{W}) \neq V(H(g))$, then we would have violated Theorem 10.3.6(1) instead.

The reason that behavior as in Example 10.3.7 is impossible in complex projective space is that the algebraic closure of the complex numbers guarantees that equations have enough solutions to prevent the kind of isolated points that we found in our example. Over the complex numbers the projective variety defined by $g(X,Y)$ will be a pinched sphere and what was our isolated point over the reals will be the pinched point which is no longer isolated (Exercise 10.3.5).

We shall have more to say about the projective completion of a variety at the end of Section 10.8 once we have a little more algebra behind us.

## 10.4 Resultants

This section introduces an important tool for determining if polynomials in one variable have a common factor. It will be needed later in several contexts.

**Definition.** Let

$$\begin{aligned} f(X) &= a_mX^m + a_{m-1}X^{m-1} + \ldots + a_0, \\ g(X) &= b_nX^n + b_{n-1}X^{n-1} + \ldots + b_0 \end{aligned} \tag{10.24}$$

be two polynomials in $D[X]$ of **positive** degrees m and n, respectively, where D is a unique factorization domain. Define the *Sylvester matrix* $SM(f,g)$ of f and g by

$$SM(f,g) = \begin{pmatrix} a_m & a_{m-1} & & \cdots & & & & \cdots & a_1 & a_0 & 0 & \cdots & 0 \\ 0 & a_m & a_{m-1} & \cdots & & & & \cdots & & a_1 & a_0 & \cdots & 0 \\ & & & \ddots & & & & \cdots & & & & \ddots & \vdots \\ 0 & & & \cdots & a_m & a_{m-1} & & \cdots & & & & \cdots & a_0 \\ b_n & b_{n-1} & & \cdots & b_1 & b_0 & 0 & \cdots & & & & \cdots & 0 \\ 0 & b_n & b_{n-1} & \cdots & & b_1 & b_0 & \cdots & & & & \cdots & 0 \\ \vdots & & & \ddots & & & & \ddots & & & & \cdots & \vdots \\ 0 & & & \cdots & & & & \cdots & b_n & b_{n-1} & & \cdots & b_0 \end{pmatrix} \tag{10.25}$$

where we have n rows of a's and m rows of b's. The determinant of SM(f,g) is called the *(Sylvester) resultant* of f and g and is denoted by R(f,g) (or $R_X(f,g)$ if we want to emphasize the fact that f and g are polynomials in X in case $a_i$ and $b_j$ are themselves polynomials in some other variables).

Note that the resultant is not a symmetric function, but it is easy to show from basic properties of the determinant that $R(g,f) = (-1)^{mn}R(f,g)$.

**10.4.1. Lemma.** If R = R(f,g) is the resultant of two polynomials f(X) and g(X) of positive degree m and n, respectively, then

$$R = G(X)f(X) + F(X)g(X),$$

where F(X) and G(X) are polynomials with deg (F) < m and deg (G) < n.

**Proof.** For each i, $1 \le i \le m + n - 1$, multiply the ith column of the Sylvester matrix (10.25) by $X^{m+n-i}$ and add the result to the last column. This will produce the matrix

$$\begin{pmatrix} a_m & a_{m-1} & & \cdots & & & & \cdots & a_1 & a_0 & 0 & \cdots & X^{n-1}f(X) \\ 0 & a_m & a_{m-1} & \cdots & & & & \cdots & & a_1 & a_0 & \cdots & X^{n-2}f(X) \\ & & & \ddots & & & & \cdots & & & & \ddots & \vdots \\ 0 & & & \cdots & a_m & a_{m-1} & & \cdots & & & & \cdots & f(X) \\ b_n & b_{n-1} & & \cdots & b_1 & b_0 & 0 & \cdots & & & & \cdots & X^{m-1}g(X) \\ 0 & b_n & b_{n-1} & \cdots & & b_1 & b_0 & \cdots & & & & \cdots & X^{m-2}g(X) \\ \vdots & & & \ddots & & & & \ddots & & & & \cdots & \vdots \\ 0 & & & \cdots & & & & \cdots & b_n & b_{n-1} & & \cdots & g(X) \end{pmatrix}. \tag{10.26}$$

Let $C_{i,j}$ denote the cofactors of the matrix in (10.26). Since both of the matrices (10.25) and (10.26) have the same determinant, expanding the determinant of (10.26) by the last column gives that

$$\begin{aligned} R(f,g) &= X^{n-1}f(X)C_{1,m+n} + X^{n-2}f(X)C_{2,m+n} + \ldots + f(X)C_{n,m+n} \\ &\quad + X^{m-1}g(X)C_{n+1,m+n} + X^{m-2}g(X)C_{n+2,m+n} + \ldots + g(X)C_{m+n,m+n} \\ &= G(X)f(X) + F(X)g(X). \end{aligned}$$

Since the polynomials F(X) and G(X) have the desired properties, Lemma 10.4.1 is proved.

**10.4.2. Lemma.** If R = R(f,g) is the resultant of two polynomials f(X) and g(X) of positive degree m and n, respectively, then R = 0 if and only if there exist **nonzero** polynomials F(X) and G(X) with deg (F) < m and deg (G) < n, so that

$$G(X)f(X) + F(X)g(X) = 0. \tag{10.27}$$

**Proof.** If R = 0, then Lemma 10.4.1 shows that there are polynomials F(X) and G(X) of degree less than m and n, respectively, satisfying equation (10.27), but we still need to show that they are nonzero. Let

$$F(X) = c_{m-1}X^{m-1} + c_{m-2}X^{m-2} + \ldots + c_0$$

and

$$G(X) = d_{n-1}X^{n-1} + d_{n-2}X^{n-2} + \ldots + d_0.$$

Collecting coefficients of powers of X in (10.27) and setting them equal to 0 gives the following system of equations:

$$\begin{aligned}
&d_{n-1}a_m + c_{m-1}b_n &&= 0\\
&d_{n-1}a_{m-1} + d_{n-2}a_m + c_{m-1}b_{n-1} + c_{m-2}b_n &&= 0\\
&\ldots\\
&d_1a_0 + d_0a_1 + c_1b_0 + c_0b_1 &&= 0\\
&d_0a_0 + c_0b_0 &&= 0.
\end{aligned} \tag{10.28}$$

This system is equivalent to the equation

$$[d_{n-1}\ d_{n-2} \ldots d_0\ c_{m-1}\ c_{m-2} \ldots\ c_0]\, SM(f,g)^T = 0. \tag{10.29}$$

Since the resultant R, which is the determinant of SM(f,g), is zero, there is a non-trivial solution and we are done.

Next, we prove the converse. Assume therefore that the nonzero polynomials F(X) and G(X) exist. We are again led to equation (10.29). Since we have a non-trivial solution, the determinant of SM(f,g), namely, R, must be 0. Lemma 10.4.2 is proved.

**10.4.3. Theorem.** Two nonconstant polynomials f(X) and g(X) have a nonconstant common factor if and only if R(f,g) = 0.

**Proof.** Let d(X) be the greatest common divisor of f(X) and g(X). It suffices to show that R = R(f,g) = 0 if and only if d(X) is a nonconstant polynomial.

First, assume that d(X) is a nonconstant polynomial. Then f(X) = d(X)H(X) and g(X) = d(X)G(X). It follows that F(X) = –H(X) and G(X) satisfy equation (10.27) in Lemma 10.4.2, and hence R = 0. Conversely, assume that R = 0. By Lemma 10.4.2, there exist nonzero polynomials F(X) and G(X) satisfying

$$G(X)f(X) + F(X)g(X) = 0$$

having degree less than m and n, respectively. It follows that f(X) divides g(X)F(X). Since f(X) is not a constant polynomial, deg F < deg f, and g(X) is not the zero polynomial, some prime factor of f must divide g(X). In other words, f(X) and g(X) have

a common factor and their greatest common divisor is not constant. Theorem 10.4.3 is proved.

**10.4.4. Corollary.** Two polynomials f(X) and g(X) defined over an algebraically closed field have a common root if and only if R(f,g) = 0.

**Proof.** Obvious.

The resultant can be used to check for multiple roots of a polynomial.

**10.4.5. Corollary.** A polynomial f(X) has a nonconstant factor of multiplicity larger than 1 if and only if R(f,f′) = 0. In particular, f(X) has a multiple root if and only if R(f,f′) = 0.

**Proof.** By Theorem B.8.10, f(X) has a nonconstant factor of multiplicity larger than 1 if and only if that factor also divides f′(X). Now use Theorem 10.4.3.

**Definition.** Let f(X) be a nonconstant polynomial. The resultant R(f,f′) is called the *resultant of f*.

Here are two useful formulas for resultants that are worth stating explicitly.

**10.4.6. Example.** The resultant of the polynomials $a_1X + a_0$ and $b_1X + b_0$ is

$$\begin{vmatrix} a_1 & a_0 \\ b_1 & b_0 \end{vmatrix}.$$

**10.4.7. Example.** The resultant of the polynomials $a_2X^2 + a_1X + a_0$ and $b_2X^2 + b_1X + b_0$ is

$$\begin{vmatrix} a_2 & a_1 & a_0 & 0 \\ 0 & a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 & 0 \\ 0 & b_2 & b_1 & b_0 \end{vmatrix} = -\begin{vmatrix} a_2 & a_1 & a_0 & 0 \\ b_2 & b_1 & b_0 & 0 \\ 0 & a_2 & a_1 & a_0 \\ 0 & b_2 & b_1 & b_0 \end{vmatrix} = \begin{vmatrix} a_2 & a_0 \\ b_2 & b_0 \end{vmatrix}^2 - \begin{vmatrix} a_2 & a_1 \\ b_2 & b_1 \end{vmatrix}\begin{vmatrix} a_1 & a_0 \\ b_1 & b_0 \end{vmatrix}.$$

**10.4.8. Theorem.** Let

$$f(X) = a_mX^m + a_{m-1}X^{m-1} + \ldots + a_0,$$
$$g(X) = b_nX^n + b_{n-1}X^{n-1} + \ldots + b_0$$

be two polynomials where the $a_i$ and $b_j$ are homogeneous polynomials of degree $m - i$ and $n - j$, respectively, in the variables $X_1, X_2, \ldots, X_r$ and $a_mb_n \neq 0$. Then the resultant $R(X_1,X_2, \ldots ,X_r)$ of f and g (with respect to X) is either identically equal to 0 or a homogeneous polynomial of degree mn.

**Proof.** The resultant, or determinant of the Sylvester matrix, consists of a sum of signed terms, each one of which is an (m + n)-fold product of factors. If the term is nonzero, then n factors are $a_i$'s and m factors are $b_j$'s. Since the degree of a nonzero (i,j)th element in the first n rows is j − i and the degree of a nonzero (s,t)th element in the last m rows is t − s, the total degree of the term is a sum of the form

$$\sum_{j\in A}\sum_{i=1}^{n}(j-i)+\sum_{j\in B}\sum_{s=1}^{m}(t-s),$$

where the sets of indices A and B are a partition of {1,2, . . . ,m + n}. This means that the total degree is

$$\frac{1}{2}(m+n)(m+n+1)-\frac{1}{2}m(m+1)-\frac{1}{2}n(n+1)=mn.$$

**10.4.9. Theorem.** Let R be the resultant of the two polynomials

$$f(X)=a_m(X-r_1)(X-r_2)\ldots(X-r_m),$$
$$g(X)=b_n(X-s_1)(X-s_2)\ldots(X-s_n),$$

where m, n > 0. Then

$$R=a_m^n b_n^m\prod_{i=1}^{m}\prod_{j=1}^{n}(r_i-s_j) \tag{10.30}$$

$$=a_m^n\prod_{i=1}^{m}g(r_i) \tag{10.31}$$

$$=(-1)^{mn}b_n^m\prod_{j=1}^{n}f(s_j). \tag{10.32}$$

**Proof.** Let us assume that $a_m$, $b_n$, $r_i$, and $s_j$ are indeterminates. Expressing f(X) and g(X) as in (10.24) we see that

$$a_i=(-1)^i a_m\sigma_i,\quad 1\le i\le m,$$
$$b_j=(-1)^j b_n\tau_j,\quad 1\le j\le n, \tag{10.33}$$

where $\sigma_i$ and $\tau_j$ are elementary symmetric polynomials in the $r_i$'s and $s_j$'s and of degree n and m, respectively. From this we see that

**Fact 1.** $a_m^n b_n^m$ divides R.

Next, we show

**Fact 2.** $r_i - s_j$ divides R for $1 \le i \le m$ and $1 \le j \le n$.

Observe that if we replace $r_i$ by $s_j$ then f(X) and g(X) have a common factor $X - s_j$. By Theorem 10.4.3, R vanishes after this substitution and hence Fact 2 is proved.

Fact 2 and the fact that the ring of polynomials $k[a_m,b_n,r_1, \ldots ,r_m,s_1, \ldots ,s_n]$ over any field k is a prime factorization domain imply that the right-hand side of equation (10.30) divides R. On the other hand, (10.33) implies that R is homogeneous of degree mn with respect to $r_i$'s and $s_j$'s. Therefore,

$$R = c\, a_m^n b_n^m \prod_{i=1}^{m}\prod_{j=1}^{n}(r_i - s_j) \tag{10.34}$$

for some constant c. To determine c, set all the $r_i$ to 0. In that case, $a_1 = a_2 = \ldots = a_m = 0$ and so the determinant of the Sylvester matrix (10.25), namely, R, is $a_m^n b_0^m$. This fact and (10.34) implies that

$$a_m^n b_0^m = c\, a_m^n b_n^m \left(\prod_{j=1}^{n}(-s_j)\right)^m . \tag{10.35}$$

The right-hand side of (10.35) is just

$$(-1)^{mn} c\, a_m^n b_n^m s_j^m = c\, a_m^n b_0^m .$$

Therefore, c = 1 and (10.30) is proved. Equalities (10.31) and (10.32) follow easily, and so the theorem is proved.

Theorem 10.4.8 leads to another useful formula for evaluating resultants.

**10.4.10. Corollary.** Given polynomials f(X), g(X), and h(X), then the resultants satisfy the product identity

$$R(fg,h) = R(f,h)\, R(g,h).$$

**Proof.** Exercise 10.4.2. For an alternate proof see [Seid68].

The resultant we have defined above addresses the problem of finding the common zeros of two polynomials. There are times when one is interested in the common zeros of a larger collection of polynomials. For properties and applications of the corresponding *multipolynomial resultant* see, for example, [CoLO98].

## 10.5 More Polynomial Preliminaries

Polynomials are the glue which holds the various aspects of algebraic geometry together. This section summarizes some additional important basic facts about polynomials and the hypersurfaces they define.

Now, we saw in Section 10.2 that if one wanted to understand curves it was important to look at these in projective space (in fact, complex projective space) since one

needs the points "at infinity." To do this one must express curves in homogeneous coordinates and via homogeneous polynomials. This means that it is important that one can relate properties of polynomials with their homogenized versions and vice versa. The next two results basically say that the answers to questions about factorization of polynomials for varieties are the same for projective varieties and their affine counterparts and so we may permit ourselves to not explicitly state which type of variety we are talking about in such cases.

**10.5.1. Proposition.** Any factor of a homogeneous polynomial is homogeneous.

**Proof.** Exercise.

**10.5.2. Theorem.** Let F be a homogeneous polynomial and f = D(F) its dehomogenization.

(1) Each factor of F dehomogenizes to a factor of f and conversely each factor of f homogenizes to a factor of F.
(2) F is irreducible if and only if f is.

**Proof.** Exercise.

**10.5.3. Proposition.** Let k be an algebraically closed field and let f(X,Y) be a homogeneous polynomial of degree d in k[X,Y]. Then f can be factored into linear factors, that is, f has the form

$$f(X,Y) = \prod_{i=1}^{d} (a_i Y - b_i X), \quad a_i, b_i \in k.$$

**Proof.** Write f in the form

$$f(X,Y) = X^d g\left(\frac{Y}{X}\right),$$

where g(Z) is a polynomial of degree d in k[Z]. Since k is algebraically closed, g (as a polynomial in Z) factors into linear factors. Replacing Z by Y/X in this factorization and simplifying gives the result.

For example,

$$\begin{aligned} X^2 - 3XY + 2Y^2 &= X^2\left[1 - 3\frac{Y}{X} + 2\left(\frac{Y}{X}\right)^2\right] \\ &= X^2\left(\frac{Y}{X} - 1\right)\left(2\frac{Y}{X} - 1\right) \\ &= (Y - X)(2Y - X). \end{aligned}$$

Note that we could equally well have put things in terms of X/Y rather than Y/X, because

$$\begin{aligned} X^2 - 3XY + 2Y^2 &= Y^2\left[\left(\frac{X}{Y}\right)^2 - 3\frac{X}{Y} + 2\right] \\ &= Y^2\left(\frac{X}{Y} - 1\right)\left(\frac{Y}{X} - 2\right) \\ &= (X - Y)(X - 2Y). \end{aligned}$$

**10.5.4. Proposition.** Let $(a_1, a_2, \ldots, a_n) \in k^n$. Every polynomial $f \in k[X_1, X_2, \ldots, X_n]$ can be written in the form

$$f = b + (X_1 - a_1)f_1 + \ldots + (X_n - a_n)f_n,$$

where $b \in k$, $f_i \in k[X_1, X_2, \ldots, X_n]$.

**Proof.** The proposition follows from the fact that

$$k[X_1, X_2, \ldots, X_n] = k[X_1 - a_1, X_2 - a_2, \ldots X_n - a_n],$$

so that $k[X_1, X_2, \ldots, X_n]$ can be thought of as a polynomial ring in $X_1 - a_1, \ldots, X_n - a_n$.

**10.5.5. Theorem.** If k is an algebraically closed field, then any nonconstant polynomial f in $k[X_1, \ldots, X_n]$, $n > 1$, has an infinite number of zeros.

**Proof.** We shall sketch a proof for the case $n = 2$ and leave the general case to the reader. Since f is not constant, we may assume without loss of generality that $f(X,Y)$ is of the form

$$f(X, Y) = a_r(Y)X^r + a_{r-1}(Y)X^{r-1} + \ldots + a_0(Y), \quad a_i(Y) \in k[Y],$$

where $r > 0$ and $a_r(Y) \neq 0$. If the degree of $a_r(Y)$ is s, then $a_r(Y)$ has at most s roots. Since k is infinite, there are infinitely many c in k so that $a_r(c) \neq 0$. Consider

$$a_r(c)X^r + a_{r-1}(c)X^{r-1} + \ldots + a_0(c) = 0.$$

This equation has a root b because k is algebraically closed. By definition, $f(b,c) = 0$. To finish the proof one simply needs to show that the infinite number of choices for c lead to an infinite number of solutions to $f(X,Y) = 0$.

**10.5.6. Theorem.** Let k be an arbitrary field, f an irreducible polynomial in $k[X_1, \ldots, X_n]$, and g an arbitrary polynomial in $k[X_1, \ldots, X_n]$. If f does not divide g, then there are only a finite number of solutions to the equations

$$f(X_1, \ldots, X_n) = g(X_1, \ldots, X_n) = 0.$$

**Proof.** Again, we shall only give a proof for the case $n = 2$ and leave the general case to the reader. By hypothesis f cannot be constant and so we may assume without loss

of generality that f(X,Y) contains a positive power of X. Consider f to belong to k(Y)[X].

**Claim.** f is irreducible in k(Y)[X].

To prove the claim, suppose that f factors in k(Y)[X], that is, $f = f_1f_2$, where $f_i \in k(Y)[X]$. The coefficients of the $f_i$ are rational functions in Y. That means that there is some polynomial a(Y) (we could use the product of all the denominators) so that $af = h_1h_2$, $h_i \in k[Y][X] = k[X,Y]$. This easily contradicts the irreducibility of f, and so the claim is proved.

Now, as polynomials in X over the field k(Y), f and g are relatively prime. It follows that there are polynomials $u_1, u_2 \in k(Y)[X]$, so that $u_1f + u_2g = 1$. Again, multiplying through by an appropriate polynomial in Y, this equation can be transformed into an equation $v_1f + v_2g = w$, where $v_1, v_2 \in k[X,Y]$ and $w(Y) \in k[Y]$. If f(a,b) = g(a,b) = 0, then w(b) = 0. But w(Y) has only a finite number of roots. For each root b of w, consider f(X,b) = 0. This is either identically equal to zero or has only a finite number of roots. It cannot be identically equal to zero, because if it were, then Y – b would divide f, which is impossible. This clearly proves that f and g have only a finite number of roots in common and finishes the proof of Theorem 10.5.6 for the case n = 2.

Theorem 10.5.6 implies a special case of the Hilbert Nullstellensatz, which is proved in Section 10.8.

**10.5.7. Corollary.** Let k be an algebraically closed field, let f be an irreducible polynomial in $k[X_1, \ldots, X_n]$, and let g be an arbitrary polynomial in $k[X_1, \ldots, X_n]$. If g vanishes wherever f does, then f divides g.

**Proof.** First, assume that f has degree 0, that is, $f \in k$. If $f \neq 0$, then the result is obvious since f is invertible. If f = 0, then g vanishes everywhere and by Theorem B.11.12 must be 0.

Now assume that f has degree r, r > 0. Since k is algebraically closed, f and g have an infinite number of common zeros by Theorem 10.5.5. If f did not divide g, then we would have a contradiction to Theorem 10.5.6.

**10.5.8. Theorem.** Consider a hypersurface V(f) (either affine or projective) defined by a polynomial f in $\mathbf{C}[X_1, \ldots, X_n]$ of the form

$$f = f_1^{n_1} f_2^{n_2} \ldots f_k^{n_k}, \tag{10.36}$$

where the $f_i$ are irreducible and nonassociates and $n_i > 0$. If V(f) = V(g) for some polynomial g in $\mathbf{C}[X_1, \ldots, X_n]$, then g has the form

$$g = c\, f_1^{m_1} f_2^{m_2} \ldots f_k^{m_k},$$

where $m_i > 0$ and c is a constant.

**Proof.** Write

$$g = g_1^{m_1} g_2^{m_2} \cdot \cdot \cdot g_s^{m_s}.$$

where the $g_j$ are irreducible and nonassociates and $m_j > 0$. Since each $f_i$ is irreducible and $f_i$ vanishes on V(g), it follows from Corollary 10.5.7 that $f_i$ divides g, that is, $g = f_i h$. Factoring h into irreducible polynomials and using the fact that we are in a unique factorization domain, $f_i$ must be an associate of some $g_j$. A similar argument shows that each $g_j$ is an associate of some $f_i$. These facts clearly imply the result.

**Definition.** Let **S** be a hypersurface in $\mathbf{C}^n$. By definition **S** = V(f) where f is a polynomial of the form shown in equation (10.36). The polynomial $f_1 f_2 \ldots f_k$ is called "the" *minimal polynomial* associated to V(f), its degree is called the *order* or *degree* of the hypersurface **S**, and the equation

$$f_1 f_2 \ldots f_k = 0$$

is called "the" *minimal equation* for **S**. The degree of a hypersurface **S** will be denoted by deg **S**.

In the context of higher dimensions the term "degree" is the one usually used, but for curves, the term "order" seems to be the more common one. It suggests more that we are talking about an invariant associated to a geometric set and not to a particular polynomial.

**10.5.9. Corollary.** Two minimal polynomials associated to a hypersurface in $\mathbf{C}^n$ differ by a nonzero constant. Two minimal equations for such a hypersurface differ by a nonzero constant.

**Proof.** This is an easy consequence of Theorem 10.5.8.

It follows from Corollary 10.5.9 that minimal polynomials and equations for hypersurfaces (over the complex numbers) are essentially unique and so in the future we are justified in referring to "the" minimal polynomial or equation. They are polynomials, respectively, equations of minimal degree. The order or degree of a hypersurface is also well defined.

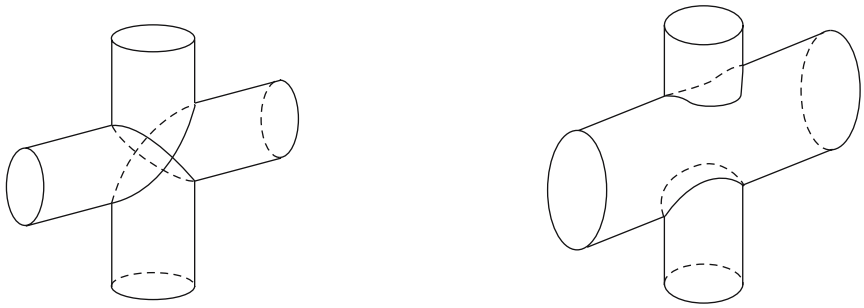
**Definition.** A variety **V** is said to be *reducible* if it is the proper union of two other varieties, that is, $\mathbf{V} = \mathbf{V}_1 \cup \mathbf{V}_2$, where $\mathbf{V}_i \neq \mathbf{V}$. Otherwise, **V** is said to be *irreducible*.

**10.5.10. Example.** In $\mathbf{R}^3$, the y-z plane V(X) and the x-z plane V(Y) are irreducible. The variety V(XY) = V(X) ∪ V(Y) is reducible.

**10.5.11. Example.** The variety V(XZ,YZ) in $\mathbf{R}^3$ is reducible because it is the union of two varieties, the z-axis and the xy-plane, that is, V(XZ,YZ) = V(Z) ∪ V(X,Y).

The next example is less trivial.

**10.5.12. Example.** Consider the varieties $\mathbf{V} = V(f_1,f_2)$ and $\mathbf{W} = V(f_1,f_3)$ in $\mathbf{R}^3$, where



(a) (b)

**Figure 10.8.** Reducible and irreducible varieties.

$$f_1(X,Y,Z) = X^2 + Y^2 - 1,$$
$$f_2(X,Y,Z) = X^2 + Z^2 - 1, \quad \text{and}$$
$$f_3(X,Y,Z) = X^2 + Z^2 - 4.$$

**V** is the intersection of two cylinders $V(f_1)$ and $V(f_2)$ of radius 1 centered on the z- and y-axis, respectively. See Figure 10.8(a). It is easy to check that **V** is the union of two ellipses that are the intersection of the cylinder $V(f_1)$ and the planes $V(X - Z)$ and $V(X + Z)$, respectively. In other words,

$$\mathbf{V} = V(f_1, X - Z) \cup V(f_1, X + Z),$$

and so **V** is reducible. The variety **W**, on the other hand, is the intersection of the cylinder $V(f_1)$ and the cylinder $V(f_3)$ of radius 2, which is centered on the y-axis. See Figure 10.8(b). It is irreducible because this time the two connected pieces of **W** cannot be separated by varieties. Although we do not yet have the tools to prove this, we sketch the steps in the argument. The reader should return to this example after reading Section 10.18. First of all, one needs to look at this as a problem in $\mathbf{C}^3$. One also needs to know about pure dimensional varieties, their degree, and Bézout's Theorem for them. The degree of an n-dimensional variety in $\mathbf{C}^m$ has to do with the number of points in the intersection of $(m - n)$-dimensional planes with the variety. The complex varieties defined by $f_1$ and $f_3$ are pure two-dimensional varieties of degree 2 (seen in $\mathbf{R}^3$ by the fact that "most" lines intersect $V(f_1)$ and $V(f_3)$ in 2 points). It follows from Bézout's Theorem (Theorem 10.18.16) that the degree of their intersection is the product of the two degrees, namely, 4. Assume that the intersection were reducible and a union of two varieties of degree r and s, respectively. Then each would contain precisely one of the two components of **W**. Such a union would have degree $r + s$. It would follow that $r + s = 4$. But this leads to a contradiction because neither degree 1 or 2 is possible (again seen in $\mathbf{R}^3$ by the fact that there are too many horizontal and almost horizontal planes that meet the curves in **W** in 4 points, something that would violate Theorem 10.18.7 and the definition of degree).

Example 10.5.12 shows that determining whether or not a general variety is irreducible can be tricky. Making this determination in the case of hypersurfaces is easier.

**10.5.13. Theorem.**

(1) A hypersurface in $\mathbf{C}^n$ with minimal equation $f = 0$ is irreducible if and only if f is irreducible.
(2) Every hypersurface in $\mathbf{C}^n$ is the union of irreducible hypersurfaces in essentially a unique way.

**Proof.** To prove (1), assume that V(f) is irreducible. Assume that f has minimal degree and

$$f = f_1 f_2 \ldots f_k,$$

where the $f_i$ are irreducible and $k > 1$. If $g = f_2 f_3 \ldots f_k$, then $f = f_1 g$ and

$$V(f) = V(f_1) \cup V(g).$$

If $V(g) \subseteq V(f_1)$, then $f_1$ would vanish on V(g). This would imply that $f_1$ divides g, which is a contradiction. Therefore, $V(g) \not\subset V(f_1)$. A similar argument shows that $V(f_1) \not\subset V(g)$. But we now have a contradiction to the fact that V(f) is irreducible.

Next, assume that f is irreducible and that $V(f) = \mathbf{V}_1 \cup \mathbf{V}_2$. Assume that $\mathbf{V}_1 \neq V(f)$. We shall prove that $\mathbf{V}_2 = V(f)$, which will show that V(f) is not reducible. Let $\mathbf{V}_1 = V(\{g_i\})$ and $\mathbf{V}_2 = V(\{h_i\})$. Now one of the $g_i$, say $g_1$, does not vanish on V(f); otherwise, $V(f) \subseteq \mathbf{V}_1$ and it would follow that $\mathbf{V}_1 = V(f)$. But $g_1 h_j$ vanishes on V(f) for all j. By Corollary 10.5.7, f divides $g_1 h_j$. Since f is irreducible and does not divide $g_1$, it must divide $h_j$. It follows that $V(f) \subseteq \mathbf{V}_2$. Clearly, $\mathbf{V}_2 \subseteq V(f)$ and so $\mathbf{V}_2 = V(f)$.

To prove (2), assume that the hypersurface $\mathbf{S}$ is defined by the minimal equation $f = 0$, where

$$f = f_1 f_2 \ldots f_k,$$

the $f_i$ are irreducible, and $k > 1$. If $\mathbf{S}_i = V(f_i)$, then the $\mathbf{S}_i$ are irreducible and distinct. Furthermore,

$$\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2 \cup \ldots \cup \mathbf{S}_k.$$

To prove that the $\mathbf{S}_i$ are unique, assume that

$$\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2 \cup \ldots \cup \mathbf{S}_k = \mathbf{T}_1 \cup \mathbf{T}_2 \cup \ldots \cup \mathbf{T}_m,$$

where the $\mathbf{T}_j$ are irreducible hypersurfaces and distinct. If $\mathbf{T}_j = V(g_j)$, where $g_j$ are irreducible polynomials, then each $g_j$ must be an associate of some $f_i$ by Theorem 10.5.8 and so $\mathbf{T}_j = \mathbf{S}_i$ and (2) is proved.

**Definition.** Let V(f) be a hypersurface defined by a polynomial f that has a factorization as shown in equation (10.36). The sets $V(f_i)$ are called the *irreducible components*, or simply *components*, of V(f).

The next result is useful when we want to localize a problem to affine space.

**10.5.14. Theorem.** Let **X** be any hypersurface in $\mathbf{P}^n(k)$ and let **p** be any point in **X**. We may always choose a coordinate system, so that the plane at infinity relative to that coordinate system is neither a component of **X** nor does it contain **p**.

**Proof.** Over **C**, there are only a finite number of planar components of **X**. Therefore, we can choose a hyperplane **Y** which is none of these and which does not contain the point **p** either. By Theorem 10.3.2 we now define a coordinate system with **Y** its plane at infinity.

What Theorems 10.3.6 and 10.5.14 allow us to do is that if we want to analyze a projective variety in $\mathbf{P}^n(\mathbf{C})$ in the neighborhood of a point we may always assume that the point and a neighborhood of it always lie in $\mathbf{C}^n$.

## 10.6 Singularities and Tangents of Plane Curves

One thing we would like to have stand out in the course of reading this chapter is the constant interplay between algebra and geometry. Seeing the geometry is especially useful when the subject matter gets very abstract, which tends to be the case in algebraic geometry. We are starting to get to some very important concepts in algebraic geometry that one can arrive at in different ways. By and large, insofar as it is possible, we emphasize a geometric approach to minimize the amount of algebraic background required of the reader. Nevertheless there are unavoidable technical aspects to definitions and theorems if we want to state things precisely; therefore, let us start with a brief overview of how our particular approach is motivated by some intimate connections between algebra and geometry.

Let **C** be an affine plane curve and f(X,Y) its minimal polynomial. In the last section we defined the degree of the **curve C** to be the degree of f(X,Y). The degree of a polynomial is a well-defined standard **algebraic** invariant associated to a polynomial. Here is a **geometric** definition of this invariant of the curve. Let

$$x = x_0 + \lambda t$$

$$y = y_0 + \mu t$$

be the parametric equations for an arbitrary line **L** through a point $\mathbf{p} = (x_0,y_0)$ on the curve. If f(X,Y) has degree d, then

$$g(t) = f(x_0 + \lambda t, y_0 + \mu t)$$

is a polynomial in t of degree $e \le d$ (e could be less than d). The roots of the polynomial g correspond to intersections of the line **L** with the curve **C**. A geometric interpretation of the degree of a plane curve is then that it is the maximum number of points that a line can intersect the curve.

How can one define tangent lines to the curve **C** at **p**? Suppose that

$$g(t) = c_k t^k + c_{k+1} t^{k+1} + \ldots + c_e t^e,$$

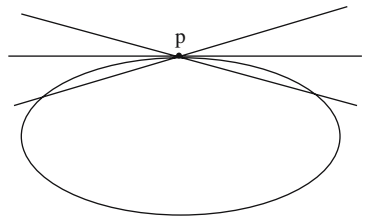



**Figure 10.9.** Comparing the tangent with non-tangent lines.

where $c_k$ and $c_e$ are nonzero. The integer k is called the *intersection multiplicity of* **C** *and* **L** *at* **p** and is denoted by $i(\mathbf{C},\mathbf{L};\mathbf{p})$. The integer $i(\mathbf{C},\mathbf{L};\mathbf{p})$ is the analog, in the context of intersections, of the multiplicity of a root of a polynomial with the same intuitive connotation. The integer k can vary from line to line, therefore, define the *multiplicity of* **C** *at* **p**, denoted by $m_{\mathbf{p}}(\mathbf{C})$, by

$$m_{\mathbf{p}}(\mathbf{C}) = \min_{\text{line } \mathbf{L} \text{ through } \mathbf{p}} i(\mathbf{C},\mathbf{L};\mathbf{p}).$$

A line **L** is then *tangent* to **C** at **p** if $i(\mathbf{C},\mathbf{L};\mathbf{p}) > m_{\mathbf{p}}(\mathbf{C})$. What this says is that a tangent line is a line that intersects the curve more often at **p** than other lines and hence in fewer points elsewhere. This agrees with one intuition of tangent lines, namely, that they are lines that intersect a curve in only one point whereas nearby lines through the point intersect the curve in additional points. See Figure 10.9.

Finally, the multiplicity $m_{\mathbf{p}}(\mathbf{C})$ defined geometrically above has an algebraic interpretation.

**Definition.** Given a polynomial $f(X_1,X_2,\ldots,X_n)$, expand f about a point $\mathbf{p} = (x_1,x_2,\ldots,x_n)$ in a finite sum of the form

$$f(X_1,X_2,\ldots,X_n) = \sum_{i_1,i_2,\ldots,i_n} a_{i_1 i_2 \ldots i_n}(X_1-x_1)^{i_1}(X_2-x_2)^{i_2}\ \ldots\ (X_n-x_n)^{i_n}.$$

The smallest degree of all the monomials appearing in the expansion above is called the *order of f at* ***p*** and is denoted by $\mathrm{ord}_{\mathbf{p}}(f)$.

It is easy to see that

$$m_{\mathbf{p}}(\mathbf{C}) = \mathrm{ord}_{\mathbf{p}}(f).$$

This concludes our overview. It should prepare the reader for the various definitions in this section and subsequent ones, in particular Section 10.17, where intersections of lines and planes with varieties are used to isolate important concepts. Right now we start back at the beginning.

Assume that **C** is a plane curve in $\mathbf{P}^2(\mathbf{C})$. Our main goal is to analyze points of **C**, in particular, certain special points, the "singular" points. In the process we shall also define tangent lines. Assume that **C** is defined by the homogeneous polynomial F(X,Y,Z) and consider a point of **C**. By Theorem 10.5.14 we can choose a coordinate system in which the point does not lie on Z = 0 and in which Z = 0 is not a component of **C**. With this choice of coordinate system we can study properties of **C** at our

point using affine coordinates. In what follows, the term "line" will refer to either an affine line or the corresponding projective line, as appropriate.

Let f(X,Y) = F(X,Y,1). The affine plane curve defined by

$$f(X,Y) = 0$$

is clearly the affine part of our projective curve **C**. Assume that $\mathbf{p} = (x_0,y_0)$ is our point of interest in affine coordinates. Let

$$\begin{aligned} x &= x_0 + \lambda t, \\ y &= y_0 + \mu t, \quad t \in \mathbf{C}, \end{aligned}$$

and

$$g(t) = f(x_0 + \lambda t, y_0 + \mu t).$$

For fixed $\lambda$ and $\mu$, g(t) can be thought of as the value of f along the line through **p** with direction vector $(\lambda,\mu)$ and parameterization $t \to (x_0 + \lambda t, y_0 + \mu t)$. The Taylor expansion for the function g(t) is

$$g(t) = g(0) + g'(0)t + \frac{1}{2!}g''(0)t^2 + \dots \tag{10.37}$$

**Definition.** The *multiplicity* of g(t) at 0 is said to be r if all the kth order derivatives of g vanish at 0 for k < r, but $g^{(r)}(0) \neq 0$.

Equation (10.37) leads to the expansion

$$\begin{aligned} f(x,y) = f(x_0,y_0) &+ (f_x(x_0,y_0)\lambda + f_y(x_0,y_0)\mu)t \\ &+ \frac{1}{2!}\left(f_{xx}(x_0,y_0)\lambda^2 + 2f_{xy}(x_0,y_0)\lambda\mu + f_{yy}(x_0,y_0)\mu^2\right)t^2 + \dots . \end{aligned} \tag{10.38}$$

But $f(x_0,y_0) = 0$, and so

$$f(x,y) = \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \frac{\partial^k f}{\partial x^i \partial y^{k-i}}(x_0,y_0)(x-x_0)^i (x-y_0)^{k-j} \tag{10.39}$$

**Definition.** The *multiplicity* of **C** (or f) at **p**, denoted by $m_{\mathbf{p}}(\mathbf{C})$ (or $m_{\mathbf{p}}(f)$), is said to be r if all the kth order partials of f vanish at **p** for k < r but at least one rth order partial of f does **not** vanish at **p**. In this case **p** is called a point of *multiplicity r*, or an *r-fold* point of **C** (or f).

**10.6.1. Proposition.** The multiplicity of a plane curve at a point is well defined and does not depend on the chosen coordinate system.

**Proof.** Since all our coordinate systems are related by linear transformations, the proposition is an easy consequence of the chain rule for derivatives.

If the multiplicity of **C** at **p** is r, then we say that every line through **p** has at least r intersections with **C** at **p**. The multiplicity tells us the minimum power of t that we can factor out of equation (10.38). On the other hand, it may be possible to factor out more powers of t depending on the line. For example, if the multiplicity is 1, then

$$f_x(x_0, y_0)\lambda + f_y(x_0, y_0)\mu = 0 \tag{10.40}$$

would imply that we can factor at least one more t out of the equation. The $\lambda$ and $\mu$ correspond to a unique line with direction vector $(\lambda,\mu)$ which is called the "tangent" line to the curve at **p**. If the multiplicity of the curve is 2 at **p**, then we can again factor another power of t out of the equation provided that

$$f_{xx}(x_0, y_0)\lambda^2 + 2f_{xy}(x_0, y_0)\lambda\mu + f_{yy}(x_0, y_0)\mu^2 = 0. \tag{10.41}$$

Equation (10.41) has up to two linearly independent root pairs $(\lambda,\mu)$. Each of these pairs is the direction vector for a line through **p**, which is again called a "tangent" line. We can continue in this way defining what is meant by "tangent" lines of higher and higher multiplicity at a point. Basically, we want to call a line with direction vector $(\lambda,\mu)$ a "tangent" line if we can factor a higher power of t out of equation (10.38) for those values of $\lambda$ and $\mu$ than is warranted by the multiplicity of the curve at that point. In the general case, the tangent lines are determined by finding the linearly independent solutions $(\lambda,\mu)$ to the equation

$$\sum_{i=0}^{r} \binom{r}{i} \frac{\partial^r f}{\partial x^i \partial y^{r-i}}(x_0, y_0)\lambda^i \mu^{r-i} = 0. \tag{10.42}$$

These observations lead to the following definition:

**Definition.** A *tangent line* to the plane curve **C** (or f) at **p** is a line through **p** with the property that if $(\lambda,\mu)$ is a direction vector for the line then $g^{(i)}(0)$ in equation (10.37) vanishes for $0 \le i \le k$, where $k > m_{\mathbf{p}}(\mathbf{C})$ (or, equivalently, g has multiplicity higher than $m_{\mathbf{p}}(\mathbf{C})$ at 0).

**10.6.2. Proposition.** Tangent lines to a plane curve at a point are well defined and do not depend on the chosen coordinate system.

**Proof.** Clear.

One can easily see that if tangent lines are counted with their multiplicities, then there are exactly $m_{\mathbf{p}}(\mathbf{C})$ tangent lines at every point on a plane curve **C** (or f).

**Definition.** The point **p** is called a *simple point* of **C** (or f) if $m_{\mathbf{p}}(\mathbf{C}) = 1$. The point **p** is called a *singular point* of **C** (or f) if $m_{\mathbf{p}}(\mathbf{C}) > 1$. A singular point is called a *double*, *triple*, etc., point if $m_{\mathbf{p}}(\mathbf{C}) = 2, 3, \ldots$, respectively. A *nonsingular* plane curve is a plane curve that has no singular points. A point of multiplicity r is called *ordinary* if the r tangents at the point are distinct.

Note that singular points of **C** (or f) are those points **p** for which

$$f(\mathbf{p}) = f_X(\mathbf{p}) = f_Y(\mathbf{p}) = 0.$$

It should also be clear that if f has no terms of degree less than r and some of degree r, then the origin is a point of multiplicity r and the tangents to **C** (or f) at the origin are the components of the equation which equates the terms of f of degree r to 0.

**10.6.3. Theorem.** If **C** and **D** are distinct irreducible curves, then every point **p** in the intersection of **C** and **D** is a singular point of $\mathbf{C} \cup \mathbf{D}$. In particular, every point that belongs to two distinct components of a curve is a singular point.

**Proof.** Let f and g be minimal polynomials for **C** and **D**, respectively. It is easy to see that fg is the minimal polynomial for $\mathbf{C} \cup \mathbf{D}$. Without loss of generality, we may assume that **p** is the origin. Then

$$\mathrm{ord}(fg) = \mathrm{ord}(f) + \mathrm{ord}(g) \geq 2,$$

which clearly implies that the origin is a singular point for fg = 0.

**10.6.4. Theorem.** A plane curve has at most a finite number of singularities.

**Proof.** First assume that the curve is irreducible and defined by the irreducible polynomial f(X,Y). Either X or Y must appear in f(X,Y). Without loss of generality assume that it is X. Then $f_X \neq 0$. Since $f_X$ has smaller degree than f, f does not divide $f_X$. It follows from Theorem 10.5.6 that $f_X$ has at most a finite number of zeros on the curve. If the curve is reducible, the result follows from the irreducible case since distinct components intersect in at most a finite number of points. Finally, the result applies to projective curves as well because a curve has only a finite number of infinite points.

**10.6.5. Theorem.** If $(x_0,y_0)$ is a simple point of a plane curve **C** defined by f(X,Y) = 0, then

$$f_X(x_0,y_0)(X - x_0) + f_Y(x_0,y_0)(Y - y_0) = 0$$

is the equation of the tangent line to **C** at $(x_0,y_0)$.

**Proof.** This is an easy consequence of equation (10.40) and the discussion above.

Below are some curves that show some of the possibilities in the behavior of a curve at a point, in this case, the origin.

**10.6.6. Example.** $X^3 - X^2 + Y^2 = 0$

**Analysis.** See Figure 10.10(a). The origin is an ordinary double point with tangent lines X + Y = 0 and X − Y = 0.

**10.6.7. Example.** $X^3 + X^2 + Y^2 = 0$

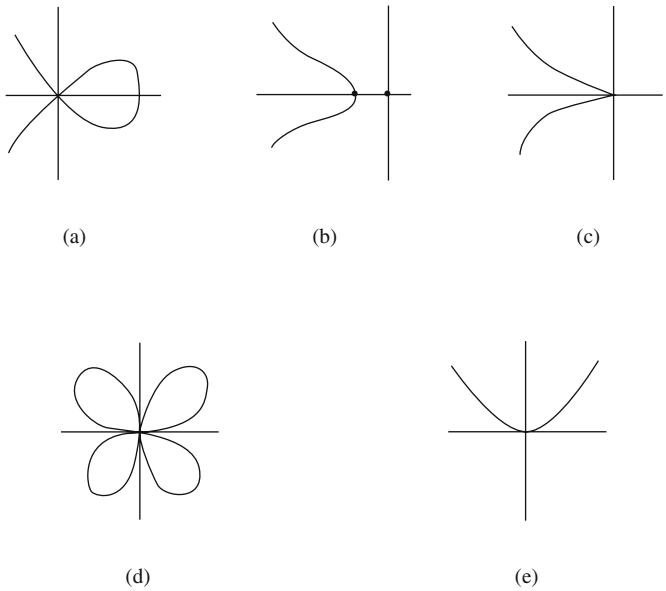



**Figure 10.10.** Different types of singularities.

**Analysis.** See Figure 10.10(b). The origin, now an isolated point of the real curve, is an ordinary double point with tangent lines $X + \mathbf{i}Y = 0$ and $X - \mathbf{i}Y = 0$.

**10.6.8. Example.** $X^3 - Y^2 = 0$

**Analysis.** See Figure 10.10(c). The origin is a double point but not ordinary. We have a "cusp."

**10.6.9. Example.** $(X^2 + Y^2)^3 - 4X^2Y^2 = 0$

**Analysis.** See Figure 10.10(d). The origin is a point of multiplicity four. There are two double tangents.

**10.6.10. Example.** $Y^6 - X^3Y^2 - X^5 = 0$

**Analysis.** See Figure 10.10(e). There is a triple tangent and two simple tangents at the origin.

The curve in example 10.6.8 is a special case of the well-known family of curves

$$X^p - Y^q = 0, \tag{10.43}$$

where $p, q > 1$. These curves capture a whole class of singularities that can be indexed by the pair of integers (p,q). In general though, singular points of plane curves are much more complicated than that and cannot be catalogued that simply. A general

singularity **can** be described in terms of a finite sequence of what are called *Puiseux pairs*! See [BriK81].

We shall show that a plane curve can have at most a finite number of singular points and that all the other points are regular. Again note that looking at real curves can be deceptive because it might seem that certain points look regular, when in fact they are not. For example, one can show that the curves defined by equation (10.43) give rise to only one of three shapes

$$X^2 - Y^3 = 0, \tag{10.44a}$$

$$X^3 - Y^4 = 0, \quad \text{or} \tag{10.44b}$$

$$X^3 - Y^5 = 0. \tag{10.44c}$$

Just like some of the plane curves in Examples 10.6.6–10.6.10, these curves may look pretty "regular" at the origin but actually have singularities there. To see this one has to deal with these curves over **C**. One can show (see [BriK81]) that the manifold that they define is not as differentiable at the singularity as it is at regular points. For example, the curve defined by equation (10.44b) is a $C^1$ submanifold of $\mathbf{R}^2$ but not a $C^2$ submanifold. When one passes to $\mathbf{C}^2$, one can show that the complex curve is never a submanifold at the singularity. In fact, singularities of complex curves are points that do not have any neighborhood **U** so that the imbedding of **U** in $\mathbf{C}^2$ looks like the imbedding of $\mathbf{R}^2$ in $\mathbf{R}^4$. More precisely, one can show that if one intersects the curve with the boundary of a ball neighborhood of the singular point in $\mathbf{C}^2$, which is homeomorphic to $\mathbf{S}^3$, one will get a knot in $\mathbf{S}^3$. In the case of our curves one gets a torus knot of type (p,q). To summarize, one can show that small neighborhoods of a singular point of a curve in $\mathbf{C}^2$ are topologically disks (assuming that they are not disconnected pieces) which are a cone on a knot in their boundary. In general, when we get disconnected pieces, neighborhoods consist of a collection of such cones.

Although we started out at the beginning of this section wanting to analyze projective plane curves, everything we have done so far was with respect to a special affine version of it. Fortunately, one can show that all the concepts we defined are independent of the particular affine representation of the curve. As an example, we rephrase the notion of multiplicity in the context of homogeneous polynomials.

**10.6.11. Theorem.** A point **p** of a plane curve **C** in $\mathbf{P}^2(\mathbf{C})$ defined in $\mathbf{P}^2(\mathbf{C})$ by the homogeneous polynomial F(X,Y,Z) has multiplicity r if and only if all the (r − 1)th derivatives of F vanish at **p**, but not all rth derivatives.

**Proof.** First of all, we may assume as before that **p** does not lie on the line Z = 0 and that Z = 0 is not a component of F. If f(X,Y) = F(X,Y,1), then f(X,Y) = 0 is the affine equation corresponding to F(X,Y,Z), that is, f(x,y) = 0 if and only if F(x,y,1) = 0. Let **p** = [a,b,1]. Since

$$\frac{\partial f}{\partial X}(a,b) = \frac{\partial F}{\partial X}(a,b,1) \quad \text{and} \quad \frac{\partial f}{\partial Y}(a,b) = \frac{\partial F}{\partial Y}(a,b,1),$$

it follows that

$$\frac{\partial f}{\partial X}(a,b) = \frac{\partial f}{\partial Y}(a,b) = 0$$

if and only if

$$\frac{\partial F}{\partial X}(a,b,1) = \frac{\partial F}{\partial Y}(a,b,1) = 0.$$

This means that there is no problem matching the zeros of the partials of f and F with respect to X and Y. The only sticky part is $\partial F/\partial Z$. However, Euler's formula (see Theorem B.7.8) implies that

$$X\frac{\partial F}{\partial X} + Y\frac{\partial F}{\partial Y} + Z\frac{\partial F}{\partial Z} = dF,$$

where d is the degree of F. Therefore,

$$F(a,b,1) = \frac{\partial F}{\partial X}(a,b,1) = \frac{\partial F}{\partial Y}(a,b,1) = 0$$

if and only if

$$\frac{\partial F}{\partial X}(a,b,1) = \frac{\partial F}{\partial Y}(a,b,1) = \frac{\partial F}{\partial Z}(a,b,1) = 0.$$

Induction and a similar argument for higher partials finish the proof of the theorem.

**10.6.12. Theorem.** If **p** is a regular point on a plane curve **C** defined by F(X,Y,Z), then the equation of the tangent to **C** at **p** is

$$X\frac{\partial F}{\partial X}(\mathbf{p}) + Y\frac{\partial F}{\partial Y}(\mathbf{p}) + Z\frac{\partial F}{\partial Z}(\mathbf{p}) = 0.$$

**Proof.** Let us use the same coordinate system and notation as in Theorem 10.6.11. Then equation (10.40) shows that a tangent line is defined by

$$f_X(a,b)(X-a) + f_Y(a,b)(Y-b) = 0,$$

and so

$$F_X(a,b,1)\left(\frac{X}{Z} - a\right) + F_Y(a,b,1)\left(\frac{Y}{Z} - b\right) = 0. \tag{10.45}$$

But since F(a,b,1) = 0, the Euler formula implies that

$$F_Z(a,b,1) = -aF_X(a,b,1) - bF_Y(a,b,1). \tag{10.46}$$

Now multiply equation (10.45) by Z and use equation (10.46) to get the desired result.

Additional facts about multiplicities and singular points are postponed to the end of the next section after we have discussed Bézout's theorem. We finish this section with an interesting fact about the intersection of lines with a curve.

**10.6.13. Theorem.** Let **C** be a plane curve order n and let **p** be a point not on **C**. Then (in $\mathbf{P}^2(\mathbf{C})$) there are at most n(n – 1) lines $\mathbf{L}_i$ through **p** so that for any other line **L** through **p**, **L** intersects **C** in **exactly** n distinct points.

**Proof.** The proof will follow the one given in [BriK81]. By a change of coordinates, we may assume that **p** = [1,0,0] and **C** is the set of zeros of the equation F(X,Y,Z) = 0 for some homogeneous polynomial F of degree n. Let $\mathbf{L}_y$ denote the line with equation X = 0. Since every line through **p** intersects $\mathbf{L}_y$ in a unique point, let us parameterize those lines by their intersection point, that is $\mathbf{L}_{s,t}$ will denote the line through **p** and the point $[0,s,t] \in \mathbf{L}_y$. By (the complex version of) Theorem 3.4.1.4 every point on $\mathbf{L}_{s,t} - \mathbf{p}$ has a unique representation [r,s,t], $r \in \mathbf{C}$. It follows that the intersections of $\mathbf{L}_{s,t}$ and the curve **C** are defined by the zeros of the equation F(r,s,t) = 0. Note that F(r,s,t) is not identically zero because $\mathbf{L}_{s,t}$ is not a component of **C**. In fact, it will have degree n in r. If we fix s and t and think of F as a polynomial in r, then by Corollary 10.4.5 F will have multiple roots for r if and only if its resultant is zero. But the resultant of F and $\partial F/\partial r$ is a homogeneous polynomial in s and t of degree n(n – 1). This can be turned into a polynomial in s/t or t/s of the same degree and hence has at most n(n – 1) zeros. Only the lines corresponding to those zeros will intersect **C** with a multiplicity higher than 1.

For example, the case n = 2 in Theorem 10.6.13 says that at most two lines through a point not on a quadratic curve in $\mathbf{C}^2$ are tangent to that curve. The case n = 3 says that at most $6 = 3 \cdot 2$ lines are tangent to a cubic. Figure 10.11 shows a cubic that actually has six such tangents.

## 10.7 Intersections of Plane Curves

We now come to one of the basic results in the theory of intersections of plane curves, namely, Bézout's theorem. This theorem is important in other areas of algebraic

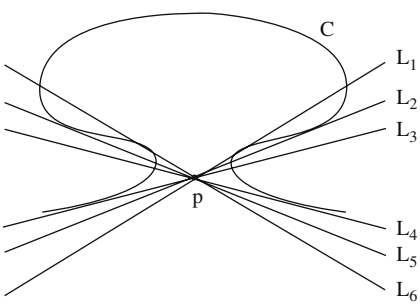



**Figure 10.11.** A cubic with six concurrent tangents.

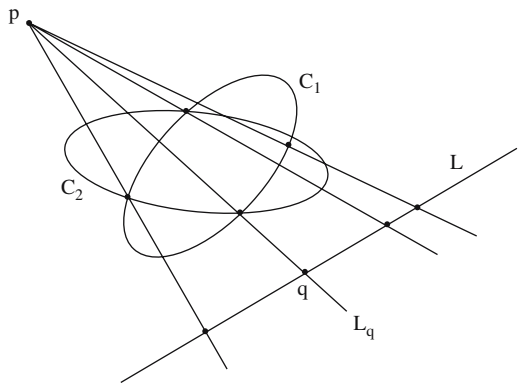



**Figure 10.12.** Defining the multiplicity of intersections.

geometry. Roughly speaking, the theorem states that if two curves in the complex plane have order m and n, respectively, then they have precisely mn points in the intersection **if** the intersection points are counted with the appropriate multiplicity. If m = 1, note the similarity between Bézout's theorem and the fundamental theorem of algebra that says that a polynomial of degree n has n roots. The latter is also only true if we have the correct notion of multiplicity of roots. Therefore part of our problem is going to be to decide how to count. We defined a notion of multiplicity of points in the intersection of a curve and a line but do not yet have a corresponding definition when we are intersecting two nonlinear curves. Different approaches to defining this notion of multiplicity exist. We shall follow the approach used in [BriK81].

Consider two plane curves $\mathbf{C}_1$ and $\mathbf{C}_2$ in $\mathbf{P}^2(\mathbf{C})$ of order m and n, respectively. The first assumption we make is that these curves have no common component (otherwise the intersection would contain an infinite number of points). Let **p** be a point which does not lie on either $\mathbf{C}_1$ or $\mathbf{C}_2$ and let **L** be any line that does not contain **p**. For each point **q** on **L** let $\mathbf{L_q}$ be the line that contains **p** and **q**. See Figure 10.12. The central projection of $\mathbf{P}^2(\mathbf{C}) - \mathbf{p}$ onto **L** maps all points of $\mathbf{L_q} - \mathbf{p}$ to **q**. Our object will now be to show two things. First, we show that only a finite number of lines through **p** contain intersection points of $\mathbf{C}_1$ and $\mathbf{C}_2$. Second, we shall define intersection multiplicities using these lines.

We begin by choosing homogeneous coordinates (X,Y,Z) for $\mathbf{P}^2(\mathbf{C})$ so that **p** has coordinates (0,0,1) and **L** is the line defined by the equation Z = 0. If we identify points (X,Y,0) on **L** with the pair of coordinates (X,Y), then the central projection of $\mathbf{P}^2(\mathbf{C}) - \mathbf{p}$ onto **L** is given by

$$(X, Y, Z) \to (X, Y).$$

Next, let $\mathbf{C}_1$ and $\mathbf{C}_2$ be defined by homogeneous polynomials F = F(X,Y,Z) and G = G(X,Y,Z) of degree m and n, respectively. Let us consider these polynomials in $\mathbf{C}[X,Y]$. Then

$$F = a_m Z^m + a_{m-1} Z^{m-1} + \ldots + a_0 \tag{10.47a}$$

and

$$G = b_n Z^n + b_{n-1} Z^{n-1} + \ldots + b_0, \tag{10.47b}$$

where $a_i$ and $b_j$ are homogeneous polynomials in $\mathbf{C}[X,Y]$ of degree $m - i$ and $n - j$, respectively. Now the fact that $a_i$ and $b_j$ are homogeneous polynomials in X and Y means that $a_i(0,0) = b_j(0,0) = 0$ whenever $i < m$ and $j < n$ and so

$$F(\mathbf{p}) = F(0,0,1) = a_m \quad \text{and} \quad G(\mathbf{p}) = G(0,0,1) = b_n.$$

Since $\mathbf{p}$ does not belong to either $\mathbf{C}_1$ or $\mathbf{C}_2$, it follows that both constants $a_m$ and $b_n$ are nonzero. Therefore, by Theorem 10.4.8, the resultant $R = R(F,G)$ of F and G thought of as polynomials in Z is either 0 or a homogeneous polynomial in X and Y of degree mn. The possibility that R is identically zero is impossible by Theorem 10.4.3 since F and G do not have any common factors. Therefore, R must be a nonzero homogeneous polynomial of degree mn.

Now assume that [x,y,z] lies on the intersection of $\mathbf{C}_1$ and $\mathbf{C}_2$. Then $F(x,y,z) = G(x,y,z) = 0$ and z is a common root of

$$a_m(x,y)z^m + a_{m-1}(x,y)z^{m-1} + \ldots + a_0(x,y) = 0 \tag{10.48a}$$

and

$$b_n(x,y)z^n + b_{n-1}(x,y)z^{n-1} + \ldots + b_0(x,y) = 0. \tag{10.48b}$$

In other words, $R(x,y) = 0$. The converse is also true, that is, if $R(x,y) = 0$, then equations (10.48) have a common root. To summarize,

**10.7.1. Lemma.** The resultant R of the polynomials F and G vanishes precisely at those points $\mathbf{q} = (x,y)$ of $\mathbf{L}$ where the lines $\mathbf{L_q}$ contain an intersection point of $\mathbf{C}_1$ and $\mathbf{C}_2$.

Our first result about the number of intersection points of two curves is the following:

**10.7.2. Theorem.** Two plane curves of order m and n, respectively, that do not have a common component have at most mn points in common.

**Proof.** Suppose that the curves have $mn + 1$ points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{mn+1}$ in common. Choose the point $\mathbf{p}$ so that $\mathbf{p}$ does not lie on any line through any of the point pairs $\mathbf{p}_i$ and $\mathbf{p}_j$, $i \neq j$. Then all the lines through $\mathbf{p}$ contain at most one intersection point $\mathbf{p}_i$. Let $\mathbf{q}_i$ be the point of $\mathbf{L}$ that is the central projection of $\mathbf{p}_i$ from $\mathbf{p}$. The $mn + 1$ points $\mathbf{q}_i$ are all distinct and roots of the resultant $R(F,G)$ (Lemma 10.7.1). This contradicts the fact that the resultant has degree mn and proves the Theorem.

We are ready to define the multiplicity of intersection points. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two plane curves in $\mathbf{P}^2(\mathbf{C})$ of order m and n, respectively, that have no common component. Let $\mathbf{p}$ be a point that does not lie on either $\mathbf{C}_1$ or $\mathbf{C}_2$ and let $\mathbf{L}$ be any line that does not contain $\mathbf{p}$. Choose homogeneous coordinates (X,Y,Z) for $\mathbf{P}^2(\mathbf{C})$ so that $\mathbf{p}$ has coordi-

nates (0,0,1) and **L** is the line defined by the equation Z = 0. Let F and G be the homogeneous polynomials which define $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively. Let R(F,G) be the resultant of F and G thought of as polynomials in Z. Theorem 10.7.2 implies that our curves $\mathbf{C}_1$ and $\mathbf{C}_2$ have at most a finite number of intersection points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$. We can therefore choose the point **p** in such a way that all the lines through **p** contain at most one intersection point. Let $\mathbf{q}_i$ be the point of **L** that is the central projection of $\mathbf{p}_i$ from **p**.

**Definition.** The *intersection multiplicity* or *order of contact* of $\mathbf{p}_i$, denoted by $m_{\mathbf{p}_i}(\mathbf{C}_1,\mathbf{C}_2)$, is defined to be the multiplicity of the zero $\mathbf{q}_i$ of the resultant R(F,G).

The definition of intersection multiplicity was with respect to a particular coordinate system. In fact, the definition is independent of this choice.

**10.7.3. Proposition.** The definition of intersection multiplicity is independent of the choice of coordinate system for the complex projective plane.

**Proof.** See [BriK81].

**10.7.4. Theorem.** (Bézout's Theorem) If $\mathbf{C}_1$ and $\mathbf{C}_2$ are two plane curves in $\mathbf{P}^2(\mathbf{C})$ of order m and n, respectively, which have no common component, then

$$\sum_{\mathbf{p}\in\mathbf{C}_1\cap\mathbf{C}_2} m_{\mathbf{p}}(\mathbf{C}_1,\mathbf{C}_2) = mn.$$

**Proof.** The theorem follows immediately from the fact that, since R(F,G) is a polynomial of degree mn, it has mn zeros when counted with their multiplicity.

A weaker form of Bézout's theorem that is an immediate consequence is

**10.7.5. Corollary.** If two plane curves of order n and m, respectively, have more than nm points in common, then they have a common component.

The next definition allows us to restate Bézout's theorem in a more suggestive manner.

**Definition.** Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two plane curves that have no component in common. The *intersection number* of $\mathbf{C}_1$ and $\mathbf{C}_2$, denoted by $\mathbf{C}_1 \bullet \mathbf{C}_2$, is defined by

$$\mathbf{C}_1 \bullet \mathbf{C}_2 = \sum_{\mathbf{p}\in\mathbf{C}_1\cap\mathbf{C}_2} m_{\mathbf{p}}(\mathbf{C}_1,\mathbf{C}_2).$$

**Restatement of Bézout's theorem** (with the same hypotheses):

$$\mathbf{C}_1 \bullet \mathbf{C}_2 = (\deg \mathbf{C}_1)(\deg \mathbf{C}_2)$$

The intersection number of two plane curves is what we needed to count the points of the intersection of two such curves correctly and with their multiplicity.

Eventually (in Section 10.17) we shall make this equation look more elegant still. If two hypersurfaces intersect transversally it will say that the degree of the intersection is equal to the product of the degrees of the hypersurfaces. We cannot do this yet because the intersection, although a variety, may not be a hypersurface (defined by a single polynomial) and we have only defined the degree of hypersurfaces so far.

The definition of intersection multiplicity as presented has the advantage that it leads relatively easily to generalizations of Bézout's theorem in higher dimensions. However, as it stands, it is difficult to work with since zeros of resultants are not easy to analyze. Therefore, the following theorem is useful. It allows one to compute the intersection multiplicities from the multiplicities of the individual curves in many cases.

**10.7.6. Theorem.** Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two plane curves that have no component in common and let $\mathbf{p}$ be an intersection point. Then

$$m_{\mathbf{p}}(\mathbf{C}_1, \mathbf{C}_2) \geq m_{\mathbf{p}}(\mathbf{C}_1) m_{\mathbf{p}}(\mathbf{C}_2)$$

with equality holding if and only if the tangents of $\mathbf{C}_1$ at $\mathbf{p}$ are disjoint from those of $\mathbf{C}_2$ at $\mathbf{p}$, that is, the curves intersect transversally.

**Proof.** See [BriK81].

Bèzout's theorem has a number of geometric applications. Among them are the theorems of Pascal and Brianchon. Another application gets an estimate of the number of singular points of a plane curve.

**10.7.7. Theorem.** Let $\mathbf{C}$ be a plane curve of order n without multiple components. Then

$$\sum_{\mathbf{p} \in \mathbf{C}} m_{\mathbf{p}}(\mathbf{C})(m_{\mathbf{p}}(\mathbf{C}) - 1) \leq n(n-1).$$

**Proof.** See [BriK81].

The result in Theorem 10.7.7 can be strengthened in the case of some special curves.

**10.7.8. Theorem.** Let $\mathbf{C}$ be an irreducible plane curve of order n. Then

$$\sum_{\mathbf{p} \in \mathbf{C}} m_{\mathbf{p}}(\mathbf{C})(m_{\mathbf{p}}(\mathbf{C}) - 1) \leq (n-1)(n-2).$$

**Proof.** See [BriK81]. The inequality cannot be improved because equality is possible as the curve

$$x^n + y^{n-1} = 0$$

shows (the origin is a point of multiplicity n − 1).

**10.7.9. Corollary.**

(1) An irreducible quadratic plane curve has no singular points.
(2) An irreducible cubic plane curve has at most one singular point and that is a double point.
(3) If an irreducible plane curve of order n has an (n – 1)-fold point then it has no other singular points.

**Proof.** These facts are a trivial consequence of Theorem 10.7.8.

## 10.8 Some Commutative Algebra

The varieties we have talked about so far were all hypersurfaces, that is, they were defined by a single polynomial. One polynomial defines a variety of codimension 1 (over **R** or **C**, depending on the context) and so the restriction might not seem unreasonable from a dimensional point of view when one studies curves in the real or complex planes. On the other hand, even in dimension two, we do not pick up all varieties in this way, because the ring of polynomials in more than one variable is not a principal ideal domain. If we want to consider varieties in dimensions higher than two, then the restriction to hypersurfaces is even more inadequate because in n-dimensional space we want to talk about sets other than (n – 1)-dimensional ones. We definitely need to allow the possibility that our space is defined by a collection of polynomials. For example, it takes two equations to define a line in 3-space. In order to be able to handle higher dimensions, we need some additional machinery. This section will give a brief overview of some fundamental results relating point sets (topology) and ideals (algebra). The algebra side of this is falls into the field of commutative algebra. In general, commutative algebra deals with commutative rings with 1. In fact, the rings are usually closely related to polynomial rings over a field or the integers.

**Definition.** Let **A** be a subset of $k^n$. Define the *ideal of* ***A***, I(**A**), by

$$I(\mathbf{A}) = \{f \in k[X_1, X_2, \ldots, X_n] \mid f(\mathbf{a}) = 0, \quad \text{for all } \mathbf{a} \in \mathbf{A}\}.$$

If **A** consists of a single point **a**, then we shall write I(**a**) rather than I(**A**).

**10.8.1. Lemma.** I(**A**) is an ideal in $k[X_1, X_2, \ldots, X_n]$.

**Proof.** Straightforward.

The map $\mathbf{A} \to I(\mathbf{A})$ associates ideals to arbitrary sets. We already have a kind of converse that associates a set of points to a finite set of polynomials, namely their zeros. It is convenient to extend this notion to arbitrary sets of polynomials.

**Definition.** Let S be a set of polynomials in $k[X_1, X_2, \ldots, X_n]$. Define a subset V(S) of $k^n$, called the *variety of S*, by

$$V(S) = \{\mathbf{p} \in k^n \mid f(\mathbf{p}) = 0 \quad \text{for all } f \in S\}.$$

**10.8.2. Lemma.** Let S be an arbitrary set of polynomials in $k[X_1, X_2, \ldots, X_n]$.

(1) $V(S) = V(\langle S \rangle)$.
(2) $V(S)$ is an algebraic variety, that is, $V(S)$ is the set of zeros of a **finite** set of polynomials.

**Proof.** Part (1) is easy. Part (2) follows from part (1) and the Hilbert Basis Theorem, which asserts that an arbitrary ideal has a finite basis.

It follows from Lemma 10.8.2 that what seemed like a new concept is actually nothing new. We could have defined a variety in the above more general way right at the beginning of the chapter (some authors do that) but we did not in order to emphasize the fact that a variety is defined by a finite set of polynomials because that finiteness property is important. In any case, nothing would have been simplified because one would have to appeal to the Hilbert Basis Theorem at some point no matter what.

At any rate, we now have correspondences

$$\text{subsets of } k^n \underset{V}{\overset{I}{\rightleftarrows}} \text{ ideals in } k[X_1, X_2, \ldots, X_n].$$

Our main goal in this section will be to analyze these two correspondences. In particular, we are interested in the following two questions:

**Question 1.** Given an ideal I of polynomials, how does its algebraic structure influence the topological structure of the set of points $V(I)$?
**Question 2.** Given a set of points $\mathbf{A}$, how does its topological structure influence the algebraic structure of the ideal $I(\mathbf{A})$?

It is clear, however, that without some restrictions on the domain of the correspondences and the field k we will not be able to say very much. For example, $V(X) = V(X^2)$. If $k = \mathbf{R}$ and $n = 1$, then $V(X^2 + 1) = V(\langle X^2 + 1 \rangle) = \phi$.

We start with some needed preliminary results. The first lists some simple properties of the I and V operators.

**10.8.3. Theorem.** Let $\mathbf{V}$, $\mathbf{V}_1$, $\mathbf{V}_2$ be varieties in $k^n$ and let I, $I_1$, $I_2$ be ideals in $k[X_1, X_2, \ldots, X_n]$.

(1) If $\mathbf{V}_1 \subseteq \mathbf{V}_2$, then $I(\mathbf{V}_2) \subseteq I(\mathbf{V}_1)$.
(2) If $I_1 \subseteq I_2$, then $V(I_2) \subseteq V(I_1)$.
(3) For all $\mathbf{V}$, $V(I(\mathbf{V})) = \mathbf{V}$. In particular, I is one-to-one on varieties.

**Proof.** The proofs of (1) and (2) are easy and left to the reader. To prove (3) we need to show two inclusions. Let $\mathbf{V} = V(f_1, f_2, \ldots, f_n)$. Let $\mathbf{I} = \langle f_1, f_2, \ldots, f_n \rangle$. Clearly, $\mathbf{V} = V(\mathbf{I})$.

**Claim 1.** $V(I(V(\mathbf{I}))) \subseteq V(\mathbf{I})$

Let $\mathbf{p} \in V(I(V(\mathbf{I})))$. Then $f(\mathbf{p}) = 0$ for all $f \in I(V(\mathbf{I}))$. Since $\mathbf{I}$ is clearly contained in $I(V(\mathbf{I}))$, $\mathbf{p}$ must belong to $V(\mathbf{I})$ and Claim 1 is proved.

**Claim 2.** $V(\mathbf{I}) \subseteq V(I(V(\mathbf{I})))$

Let $\mathbf{p} \in V(\mathbf{I})$. Given an arbitrary $f \in I(V(\mathbf{I}))$, $f(\mathbf{p}) = 0$ by definition and so Claim 2 is obvious. The two inclusions in Claim 1 and 2 prove the equality in (3).

Next, we state another of the really important theorems of algebraic geometry, the Hilbert Nullstellensatz. This theorem gets used in many places and in particular we shall need it. We state two variants. The proofs are not that difficult but still too long to present here.

**10.8.4. Theorem.** (The Hilbert Nullstellensatz: weak form) Let k be an algebraically closed field. If I is a proper ideal in $k[X_1,X_2, \ldots ,X_n]$, then there is at least one point on which all polynomials of I vanish.

**Proof.** See [CoLO97].

**10.8.5. Theorem.** (The Hilbert Nullstellensatz) Let k be an algebraically closed field and let f, $f_1$, $f_2$, . . . , and $f_k$ be polynomials in $k[X_1,X_2, \ldots ,X_n]$. If f vanishes on all the common zeros of the $f_i$, then some power of f is a linear combination of the $f_i$, that is, for some $m > 0$,

$$f^m = a_1f_1 + a_2f_2 + \ldots + a_kf_k.$$

In other words, if I is an ideal in $k[X_1,X_2, \ldots ,X_n]$ then $I(V(I)) = \sqrt{I}$.

**Proof.** See [CoLO97]. The proof uses Theorem 10.8.4.

An immediate consequence of the Nullstellensatz is that we can state the correspondence between varieties and ideals more precisely.

**10.8.6. Corollary.** If k is algebraically closed, then the maps I and V define correspondences

$$\text{varieties of } k^n \underset{V}{\overset{I}{\rightleftarrows}} \text{radical ideals in } k[X_1, X_2, \ldots, X_n].$$

that are one-to-one and onto.

**Proof.** We already know from Theorem 10.8.3(3) that $V(I(\mathbf{V})) = \mathbf{V}$. To show that $I(V(I)) = I$ for all radical ideals I we simply observe that $I(V(I)) = \sqrt{I}$ by the Nullstellensatz and $\sqrt{I} = I$ whenever I is a radical ideal.

Another consequence of the Nullstellensatz is an algebraic characterization of points.

**10.8.7. Theorem.** Every ideal in $k[X_1,X_2,\ldots,X_n]$ of the form $<X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n>$, $a_i \in k$, is maximal. To put it another way, for every point $\mathbf{p}$ in $k^n$, $I(\mathbf{p})$ is a maximal ideal.

**Proof.** Let $\mathbf{p} = (a_1,a_2,\ldots,a_n)$. Every polynomial $f \in k[X_1,X_2,\ldots,X_n]$ can be written in the form

$$f = b + (X_1 - a_1)f_1 + (X_2 - a_2)f_2 + \ldots + (X_n - a_n)f_n,$$

where $b \in k$ and $f_i \in k[X_1,X_2,\ldots,X_n]$. It follows that $f \in I(\mathbf{p})$ if and only if $b = 0$. This shows that the polynomials $X_i - a_i$ generate $I(\mathbf{p})$, that is,

$$I(\mathbf{p}) = <X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n>.$$

On the other hand, if f does not belong to $I(\mathbf{p})$, then $b \neq 0$. The ideal generated by $I(\mathbf{p})$ and f will then contain b and therefore be the whole ring $k[X_1,X_2,\ldots,X_n]$. It follows that $I(\mathbf{p})$ is maximal.

Note that $I(\mathbf{p})$ is a prime ideal. This fact is easy to prove directly but it also follows from the general fact that every maximal ideal is prime. In addition, the proof of Theorem 10.8.7 basically showed that

$$\frac{k[X_1,X_2,\ldots,X_n]}{I(\mathbf{p})}$$

is isomorphic to k. Hilbert's Nullstellensatz implies the following converse to Theorem 10.8.7:

**10.8.8. Theorem.** If k is algebraically closed, then every maximal ideal I of $k[X_1,X_2,\ldots,X_n]$ has the form $<X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n>$, for some $a_i \in k$.

**Proof.** We follow the proof given in [CoLO97]. Since $I \neq k[X_1,X_2,\ldots,X_n]$, it follows from the weak form of the Hilbert Nullstellensatz that V(I) is not empty. Assume that $\mathbf{p} = (a_1,a_2,\ldots,a_n) \in V(I)$. Clearly,

$$I(V(I) \subset I(\mathbf{p}).$$

By the strong Hilbert Nullstellensatz $I(V(I)) = \sqrt{I}$. Since I is maximal and every maximal ideal is prime, it follows that $\sqrt{I} = I$. Therefore,

$$I \subseteq I(\mathbf{p}) = <X_1 - a_1, X_2 - a_2, \ldots, X_n - a_n> \subset k[X_1,X_2,\ldots,X_n].$$

Since I is maximal, we must have $I = I(\mathbf{p})$, and the theorem is proved.

Theorems 10.8.7 and 10.8.8 imply

**10.8.9. Corollary.** If k is algebraically closed, then the map

$$k^n \to \text{set of maximal ideals in } k[X_1, X_2, \ldots, X_n]$$
$$\mathbf{p} \to I(\mathbf{p})$$

is one-to-one and onto.

What Corollary 10.8.9 shows is that the ring $k[X_1, X_2, \ldots, X_n]$ completely determines $k^n$ when k is algebraically closed. Since one can define a topology on the set of maximal ideals, $k[X_1, X_2, \ldots, X_n]$ also determines the topology of $k^n$. These facts are actually not that surprising because they reflect similar facts about general topological spaces. For example, a theorem of I. Gelfand and A. Kolmogoroff asserts that the ring of continuous functions $C(\mathbf{X})$ on a compact space $\mathbf{X}$ determines $\mathbf{X}$. Specifically, the map which sends a point $\mathbf{p} \in \mathbf{X}$ to the maximal ideal in $C(\mathbf{X})$ consisting of the functions that vanish on $\mathbf{p}$ is a homeomorphism. This theorem is false if $\mathbf{X}$ is not compact, but generalizations are known.

Next, we look at Questions 1 and 2 above with respect to general varieties and ideals. As a warm up, we show how the result in Theorem 10.5.13(1) regarding hypersurfaces translates into the language of ideals. First, given a polynomial f in k[X], write f in the form

$$f = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

where the $p_i$ are irreducible polynomials and $p_i$ and $p_j$ are nonassociates for $i \neq j$. Let $g = p_1 p_2 \ldots p_k$ be the minimal polynomial associated to V(f). Both f and g are polynomials for the same hypersurface. What is the algebraic relation between f and g?

**10.8.10. Lemma.** The ideal $\langle g \rangle$ is the radical of $\langle f \rangle$.

**Proof.** Easy.

**10.8.11. Lemma.** If f is an irreducible polynomial in k[X], then the (principal) ideal generated by f, $\langle f \rangle$, is a prime ideal.

**Proof.** Easy.

Lemma 10.8.11 generalizes to a characterization of the irreducibility of an arbitrary variety in terms of its ideal.

**10.8.12. Theorem.** An affine variety $\mathbf{V}$ in $k^n$ is irreducible if and only if $I(\mathbf{V})$ is a prime ideal.

**Proof.** First assume that $\mathbf{V}$ is irreducible and $f_1 f_2 \in I(\mathbf{V})$. Let $\mathbf{V}_i = \mathbf{V} \cap V(f_i)$. The sets $\mathbf{V}_i$ are varieties. Let $\mathbf{W} = \mathbf{V}_1 \cup \mathbf{V}_2$. Clearly, $\mathbf{W} \subset \mathbf{V}$. Furthermore, if $(f_1 f_2)(\mathbf{p}) = 0$, then $f_1(\mathbf{p}) = 0$ or $f_2(\mathbf{p}) = 0$. This easily implies that $\mathbf{V} \subset \mathbf{W}$. Therefore, $\mathbf{V} = \mathbf{W}$. But $\mathbf{V}$ is irreducible, so either $\mathbf{V} = \mathbf{V}_1$ or $\mathbf{V} = \mathbf{V}_2$. In other words, either $f_1$ or $f_2$ vanishes on $\mathbf{V}$, which means that either $f_1$ or $f_2$ belongs to $I(\mathbf{V})$. This shows that $I(\mathbf{V})$ is prime.

Conversely, assume that $I(\mathbf{V})$ is prime and $\mathbf{V} = \mathbf{V}_1 \cap \mathbf{V}_2$. Assume that $\mathbf{V} \neq \mathbf{V}_1$. We shall show that $\mathbf{V} = \mathbf{V}_2$, which would prove that $\mathbf{V}$ is irreducible.

**Claim.** $I(\mathbf{V}) = I(\mathbf{V}_2)$.

First of all, $\mathbf{V}_2 \subset \mathbf{V}$ implies that $I(\mathbf{V}) \subset I(\mathbf{V}_2)$. On the other hand, since $\mathbf{V}_1$ is contained in but not equal to $\mathbf{V}$ there must exist a polynomial $f_1 \in I(\mathbf{V}_1) - I(\mathbf{V})$. Let $f_2$ be any polynomial in $I(\mathbf{V}_2)$. It follows that $f_1f_2$ vanishes on $\mathbf{V}$, that is, $f_1f_2 \in I(\mathbf{V})$. Since $I(\mathbf{V})$ is prime, either $f_1$ or $f_2$ belongs to $I(\mathbf{V})$. It must be $f_2$ because $f_1$ does not by hypothesis. Since $f_2$ was an arbitrary element of $I(\mathbf{V}_2)$, we have shown the opposite inclusion $I(\mathbf{V}_2) \subset I(\mathbf{V})$ and the Claim is proved.

The claim and Theorem 10.8.3(3) imply that $\mathbf{V} = \mathbf{V}_2$ and the theorem is proved.

Theorem 10.8.12 and Corollary 10.8.6 lead to

**10.8.13. Corollary.** If k is algebraically closed, then the maps I and V define correspondences

$$\text{irreducible varieties of } k^n \underset{V}{\overset{I}{\rightleftarrows}} \text{prime ideals in } k[X_1, X_2, \ldots, X_n].$$

that are one-to-one and onto.

While we are on the subject of irreducibility, Theorem 10.8.14 may be an important theoretical result but not as useful as one would like, because checking the primality of an ideal is not that easy. Therefore, it is nice to have a theorem like the one below because it provides a practical means to establishing the irreducibility of many varieties.

**10.8.14. Theorem.** Let k be an infinite field. If $\mathbf{V}$ is a variety in $k^n$ that admits a parameterization of the form

$$p(t_1, t_2, \ldots, t_m) = (f_1(t_1, t_2, \ldots, t_m), f_2(t_1, t_2, \ldots, t_m), \ldots, f_n(t_1, t_2, \ldots, t_m)),$$

where $f_i \in k[t_1, t_2, \ldots, t_m]$, then $\mathbf{V}$ is irreducible.

**Proof.** Let

$$I(\mathbf{V}) = \{h \in k[X_1, X_2, \ldots, X_n] \mid h(p(t_1, t_2, \ldots, t_m)) \text{ is the zero polynomial in } k[t_1, t_2, \ldots, t_m]\}.$$

By Theorem 10.8.12, to prove the theorem in this case, it suffices to show that $I(\mathbf{V})$ is a prime ideal. Let $r, s \in k[X_1, X_2, \ldots, X_n]$ and assume that $rs \in I(\mathbf{V})$. Then $(rs)\circ p = (r\circ p)(s\circ p)$ is the zero polynomial in $k[t_1, t_2, \ldots, t_m]$ and so either $r\circ p$ or $s\circ p$ is the zero polynomial. In other words, either r or s belongs to $I(\mathbf{V})$ and we are done.

As one example of an application of Theorem 10.8.14, we see that all planes in $\mathbf{R}^n$ are irreducible because they can be parameterized by linear polynomials. Another example is a parabola in $\mathbf{R}^2$. Later we shall see (Theorem 10.14.6) that Theorem 10.8.14 generalizes to rational functions.

The next step in our program is to generalize Theorem 10.5.13(2) from hypersurfaces to arbitrary varieties.

**10.8.15. Theorem.** Every variety $\mathbf{V}$ can be expressed as a finite union of irreducible varieties.

**Proof.** If $\mathbf{V}$ is reducible, then $\mathbf{V} = \mathbf{W}_1 \cup \mathbf{V}_1$, where $\mathbf{W}_1$ and $\mathbf{V}_1$ are two proper subvarities of $\mathbf{V}$. If one of these, say $\mathbf{V}_1$ is reducible, then $\mathbf{V}_1 = \mathbf{W}_2 \cup \mathbf{V}_2$, where $\mathbf{W}_2$ and $\mathbf{V}_2$ are two proper subvarities of $\mathbf{V}_1$. If the theorem is false for some $\mathbf{V}$, then we could continue on like this to get a strictly decreasing chain $\mathbf{V} \supset \mathbf{V}_1 \supset \mathbf{V}_2 \supset \ldots$ of subvarieties $\mathbf{V}_i$. This would give a strictly increasing chain of ideals $I(\mathbf{V}) \subset I(\mathbf{V}_1) \subset I(\mathbf{V}_2) \subset \ldots$, contradicting the Hilbert Basis Theorem.

**Definition.** A union

$$\mathbf{V} = \mathbf{V}_1 \cup \mathbf{V}_2 \cup \ldots \cup \mathbf{V}_n$$

is called an *irredundant union* if no $\mathbf{V}_i$ is contained in any $\mathbf{V}_j$ for $i \neq j$.

**10.8.16. Theorem.** Let

$$\mathbf{V} = \mathbf{V}_1 \cup \mathbf{V}_2 \cup \ldots \cup \mathbf{V}_s = \mathbf{W}_1 \cup \mathbf{W}_2 \cup \ldots \cup \mathbf{W}_t,$$

where the $\mathbf{V}_i$ and $\mathbf{W}_j$ are irreducible and we have irredundant unions. Then $\{\mathbf{V}_i\} = \{\mathbf{W}_j\}$.

**Proof.** We have

$$\mathbf{V}_i = \mathbf{V}_i \cap \mathbf{V} = \mathbf{V}_i \cap \left(\bigcup_j \mathbf{W}_j\right) = \bigcup_j (\mathbf{V}_i \cap \mathbf{W}_j).$$

Since $\mathbf{V}_i$ is irreducible, we must have $\mathbf{V}_i = \mathbf{V}_i \cap \mathbf{W}_j$ for some j. In other words, $\mathbf{V}_i \subseteq \mathbf{W}_j$. A similar argument shows that $\mathbf{W}_j \subseteq \mathbf{V}_s$ for some s. The inclusions $\mathbf{V}_i \subseteq \mathbf{W}_j \subseteq \mathbf{V}_s$ and the irredundancy of the unions imply that $\mathbf{V}_i = \mathbf{W}_j$.

Theorems 10.8.15 and 10.8.16 say that every variety can be expressed as a finite irredundant union of irreducible varieties in only essentially one way. This justifies the next definition that generalizes one in Section 10.5.

**Definition.** If a variety $\mathbf{V}$ is written as a finite irredundant union of irreducible varieties $\mathbf{V}_i$, then the $\mathbf{V}_i$ are called the *irreducible components* of $\mathbf{V}$.

Do the analogs of the last two results about varieties hold for ideals? The relevant facts are:

(1) The Hilbert Basis Theorem (Theorem B.7.9) and Corollary B.7.4 imply that in a polynomial ring over a field one gets a unique factorization of ideals into primary ideals whose associated prime ideals are unique.
(2) Prime ideals are irreducible. (Irreducible ideals are not necessarily prime but they are primary in a Noetherian ring. A primary ideal is not necessarily irreducible.)

We do not quite get the unique decomposition of an arbitrary ideal into irreducible ideals (see Exercise 10.8.2), but we are close. We can prove

**10.8.17. Theorem.** If k is algebraically closed, then every radical ideal in $k[X_1, X_2, \ldots, X_n]$ can be expressed as a unique finite irredundant intersection of prime ideals.

**Proof.** The theorem follows from Corollary 10.8.6 and Theorems 10.8.15 and 10.8.16.

Finally, to press home the close relationship between algebra and geometry further, we show how some other algebraic operations on ideals correspond to set-theoretic operations on varieties and then summarize everything in Table 10.8.1. We first need one more definition to capture the algebraic analog of the difference of two varieties.

**Definition.** Let I and J be ideals in $k[X_1,X_2, \ldots ,X_n]$. Define the *ideal quotient* or *colon ideal*, $I:J$, by

$$I : J = \{p \in k[X_1, X_2, \ldots, X_n] \mid pq \in I \text{ for all } q \in J\}.$$

**10.8.18. Proposition.** The ideal quotient of two ideals I and J in $k[X_1,X_2, \ldots ,X_n]$ is an ideal that contains I.

**Proof.** Straightforward.

**10.8.18. Theorem.** Let I and J be ideals in $k[X_1,X_2, \ldots ,X_n]$.

**Table 10.8.1** The correspondence between affine algebraic and geometric concepts.

| **Algebra** | **Geometry** |
|---|---|
| radical ideals $I, J \subseteq k[X_1,X_2, \ldots , X_n]$ | varieties $V, W \subseteq k^n$ |
| (the field k is assumed to be algebraically closed) | |
| Radical ideal | Varieties |
| $I$ | $V(I)$ |
| $I(\mathbf{V})$ | $\mathbf{V}$ |
| Inclusion of ideals | Reverse inclusion of varieties |
| $I \subseteq J$ | $V(I) \supseteq V(J)$ |
| Addition of ideals | Intersection of varieties |
| $I + J$ | $V(I) \cap V(J)$ |
| $\sqrt{I(\mathbf{V}) + I(\mathbf{W})}$ | $\mathbf{V} \cap \mathbf{W}$ |
| Product of ideals | Union of varieties |
| $I\,J$ | $V(I) \cup V(J)$ |
| $\sqrt{I(\mathbf{V})\, I(\mathbf{W})}$ | $\mathbf{V} \cup \mathbf{W}$ |
| Intersection of ideals | Union of varieties |
| $I \cap J$ | $V(I) \cup V(J)$ |
| $I(\mathbf{V}) \cap I(\mathbf{W})$ | $\mathbf{V} \cup \mathbf{W}$ |
| Quotient of ideals | Difference of varieties |
| $I:J$ | $\overline{V(I) - V(J)}$ |
| $I(\mathbf{V}):I(\mathbf{W})$ | $\overline{\mathbf{V} - \mathbf{W}}$ |
| Prime ideal | Irreducible variety |
| Maximal ideal | Point of affine space |

(1) $V(I + J) = V(I) \cap V(J)$.
(2) $V(I \bullet J) = V(I) \cup V(J)$.
(3) $V(I \cap J) = V(I) \cup V(J)$.
(4) We always have $V(I:J) \supset \overline{V(I) - V(J)}$, but if k is algebraically closed and I is a radical ideal, then $V(I:J) = \overline{V(I) - V(J)}$. ($\overline{X}$ is the closure of the space X.)

**Proof.** Parts (1)–(3) are straightforward. Part (4) needs the Hilbert Nullstellensatz. See, for example, [CoLO97].

Up to here we have dealt with affine varieties in this section. We now turn our attention to projective varieties and show how one can associate ideals to them also. Recall that to define projective varieties we had to use homogeneous polynomials. Therefore, it might seem as if the natural ideals to associate to projective varieties are those that consist of homogeneous polynomials. The problem with that is that the sum of two homogeneous polynomials may not be homogeneous. The right definition is the following:

**Definition.** An ideal in $k[X_1,X_2, \ldots ,X_n]$ is said to be *homogeneous* if it is generated by a set of homogeneous polynomials.

**Definition.** If **V** is a projective variety in $\mathbf{P}^n(k)$, define

$$I(\mathbf{V}) = \langle f \in k[X_1, X_2, \ldots, X_n] \mid f \text{ is a homogeneous polynomial that is zero on } \mathbf{V} \rangle .$$

If I is a homogeneous ideal of $k[X_1,X_2, \ldots ,X_n]$, define

$$V(I) = \{\mathbf{p} \in \mathbf{P}^n(k) \mid f(\mathbf{p}) = 0 \text{ for all homogeneous polynomials } f \in I\}.$$

**10.8.20. Proposition.**

(1) An ideal I of $k[X_1,X_2, \ldots ,X_n]$ is a homogeneous ideal if and only if $I = \langle f_1,f_2, \ldots ,f_s \rangle$, where the $f_i$ are homogeneous polynomials.
(2) If I is a homogeneous ideal in $k[X_1,X_2, \ldots ,X_n]$, then $\sqrt{I}$ is a homogeneous ideal.
(3) If **V** is a projective variety in $\mathbf{P}^n(k)$, then $V(I(\mathbf{V})) = \mathbf{V}$.

**Proof.** The proofs are not hard. Part (1) needs the Hilbert Basis Theorem.

It follows from Proposition 10.8.20(1) that V(I) is a projective variety for all homogeneous ideals I. With the definitions above and Proposition 10.8.20 one can establish a correspondence between projective varieties and radical homogenous ideals over algebraically closed fields similar to what we had in the affine case. For the details and some minor required modifications see [CoLO97]. See Exercise 10.8.5 for another approach to projective varieties and their algebraic counterparts.

This seems like a good time to introduce another concept.

**Definition.** The *Zariski topology* of a variety **V** in $k^n$ or $\mathbf{P}^n(k)$ is the topology defined by saying that its closed sets are the subvarieties of **V**. Alternatively, the sets

$$\{\mathbf{p} \in \mathbf{V} \mid f(\mathbf{p}) \neq 0\},$$

where f is a polynomial in $k[X_1,X_2, \ldots ,X_n]$ that is homogeneous if **V** is a projective variety, define a base for the open sets of the Zariski topology.

The Zariski topology is a topology because arbitrary intersections of varieties are varieties by Lemma 10.8.2(2). This new topology on a variety is quite different from the topology induced by the standard topology of $k^n$ or $\mathbf{P}^n(k)$ when k = **R** or **C**. For one thing, all open sets intersect. (See Exercise 10.8.6 for some of its properties.) Nevertheless, it is convenient terminology which is often used in algebraic geometry. Varieties are sometimes called *closed sets* and some authors call an open subset of a projective variety a *quasiprojective variety*. The latter term is in an attempt to unify the concept of affine and projective variety. Projective varieties are clearly quasiprojective varieties, but affine varieties over **C** are also because of Theorem 10.3.6 on the projective completion of an affine variety. On the other hand, it turns out that the set of quasiprojective varieties is bigger than the set of affine and projective varieties. A lot of what we do in this chapter could be done in terms of quasiprojective varieties, but we shall not in order to cut down on the abstraction. At any rate, we shall run into the Zariski topology again at several places later on. From now on any topological statements about varieties will refer to it unless otherwise stated.

We finish this section by describing the algebraic analog of the projective completion.

**Definition.** If I is an ideal in $k[X_1,X_2, \ldots ,X_n]$, then

$$H(I) = \{H(f) \mid f \in I\}$$

is called the *homogenization of I*.

**10.8.21. Proposition.** If I is an ideal in $k[X_1,X_2, \ldots ,X_n]$, then, H(I) is a homogeneous ideal in $k[X_1,X_2, \ldots ,X_{n+1}]$.

**Proof.** Easy.

**Definition.** Let **V** be an (affine) variety in $k^n$. The projective variety V(H(I(**V**))) in $\mathbf{P}^n(k)$ is called the *projective closure* of **V**.

**10.8.22. Theorem.** Let **V** be an affine variety in $k^n$.

(1) The projective closure of **V** is the same as the projective completion of **V**.
(2) If **V** is irreducible, then so is the projective closure of **V**.

**Proof.** The proof is not hard. See [CoLO97].

## 10.9 Defining Parameterized Curves Implicitly

Suppose that we have a curve in the plane whose points are parameterized by rational functions, say,

$$x = \frac{p_1(t)}{q_1(t)}$$
$$y = \frac{p_2(t)}{q_2(t)}.$$

Clearing denominators gives

$$\begin{aligned} xq_1(t) - p_1(t) &= 0 \\ yq_2(t) - p_2(t) &= 0. \end{aligned} \tag{10.49}$$

If we could eliminate the variable t, then we would have an implicit equation in x and y only. For example, consider the parameterization

$$x = 2t - 1$$
$$y = 3t + 5$$

of a line **L** in the plane. If we solve both equations for t, set the results equal, and simplify, we get

$$3x - 2y + 7 = 0,$$

which is an equation for **L**. We could also find the equation for a curve **C** defined by quadratic parameterizations in this way, such as, for example,

$$x = t^2 - 3t + 1$$
$$y = 2t^2 + 7.$$

We again simply solve for t and set the results equal. However, as the degree of the parameterization got higher, solving for t would get more and more complicated. In fact, if the degree was larger than 4, then this approach would not work at all because there is a well-known theorem that states that there is no solution by radicals of the general equation of degree five or higher. Something else is needed if we want to find implicit definitions for spaces defined parametrically.

Recall resultants. What do they have to do with eliminating variables? Well, consider the equations (10.49) again and rewrite them as

$$\begin{aligned} f(t) &= xq_1(t) - p_1(t) = 0 \\ g(t) &= yq_2(t) - p_2(t) = 0. \end{aligned} \tag{10.50}$$

The polynomials f and g are now thought of as polynomials in t with coefficients in $\mathbf{R}[x,y]$. If

$$x_0 = \frac{p_1(t_0)}{q_1(t_0)}$$
$$y_0 = \frac{p_2(t_0)}{q_2(t_0)},$$

then the two equations in (10.50) have a common root $t_0$ when $x = x_0$ and $y = y_0$. It follows that

$$R(f,g) = 0.$$

The latter is an equation in x and y and corresponds to having eliminated t from the equations in (10.49).

**10.9.1. Example.** To find an implicit equation for the curve parameterized by

$$x = t^2 \tag{10.51a}$$

$$y = t^3 - t. \tag{10.51b}$$

**Solution.** If

$$f(t) = x - t^2$$
$$g(t) = y - t^3 - t,$$

then

$$R(f,g) = \begin{vmatrix} -1 & 0 & x & 0 & 0 \\ 0 & -1 & 0 & x & 0 \\ 0 & 0 & -1 & 0 & x \\ -1 & 0 & 1 & y & 0 \\ 0 & -1 & 0 & 1 & y \end{vmatrix} = -y^2 + x^3 - 2x^2 + x.$$

Of course, we could also have eliminated the t from equation (10.51a) directly, namely, $t = \sqrt{x}$, and substituted into equation (10.51b) to get

$$y = \sqrt{x}(x-1).$$

Squaring both sides of this equation gives the same equation in x and y as the one we got from the resultant.

**10.9.2. Example.** To find an implicit equation using resultants for the standard rational parameterization of the unit circle given by

$$x = \frac{1-t^2}{1+t^2}$$
$$y = \frac{2t}{1+t^2}.$$

**Solution.** Let

$$f(t) = x(1+t^2) + t^2 - 1$$
$$g(t) = y(1+t^2) - 2t.$$

Then

$$R(f,g) = 4\left(x^2 + y^2 - 1\right) = 0$$

is essentially the standard implicit definition of the circle.

**10.9.3. Example.** To find an implicit equation using resultants for the unit sphere **S** with center (0,0,1) parameterized by

$$x = \frac{2s}{1+s^2+t^2} \tag{10.52a}$$

$$y = \frac{2st}{1+s^2+t^2} \tag{10.52b}$$

$$z = \frac{2s^2}{1+s^2+t^2}. \tag{10.52c}$$

**Solution.** First, note that this sphere has an obvious implicit representation of the form

$$x^2 + y^2 + (z-1)^2 = 1,$$

which simplifies to

$$x^2 + y^2 + z^2 - 2z = 0.$$

On the other hand, if we use resultants to eliminate t first from equations (10.52a) and (10.52b) and then from equations (10.52b) and (10.52c), we get equations

$$4s^2\left(y^2 + s^2x^2 + x^2 - 2sx\right) = 0$$

and

$$4s^2\left(s^2z^2 + z^2 - 2s^2z + s^2y^2\right) = 0,$$

respectively. To eliminate the parameter s we find the resultant of the last two equations. However, to simplify the computations, we eliminate the factors $4s^2$ first. We know from the product rule given in Corollary 10.4.10 that the result will be a polynomial which divides the actual resultant but is good enough for what we want. We get

$$\left(x^2 + y^2 + z^2 - 2z\right)\left[y^6 + y^4\left(x^2 + z^2\right) - 2y^2z\left(y^2 + 2x^2\right) + 4x^2z^2\right] = 0. \tag{10.53}$$

The first factor is of course what we would want, but notice the second factor that corresponds to some extraneous values.

Example 10.9.3 points out a major problem with using the Sylvester resultant for finding implicit equations from parameterizations. One may get a much higher degree polynomial equation than is necessary and, since factoring polynomials is a nontrivial problem, there would be no easy way to eliminate the extraneous factors in general. From a computational point of view one always has to worry about numerical stability, so that it is important to get polynomials with as small degree as possible. A more efficient method for implicitizing parametric objects is needed.

Finally, in showing how resultants can be used to implicitize objects, we have assumed above that computations are carried out with exact arithmetic. Unfortunately, when one tries to do this on a computer it is easy to give examples, where, due to round-off errors, small variations in polynomial coefficients can lead to wildly different and incorrect implicit forms. Although exact rational arithmetic is possible it is very expensive computationally. For a more efficient way to deal with the round-off problem here see, for example, [Hobb91].

## 10.10 Gröbner Bases

This section will define and discuss applications of certain special bases of polynomial ideals called Gröbner bases. These bases lead to algorithms that not only solve a large variety of problems but provide efficient solutions to them. We shall use them to give answers to the following questions:

(1) Does an ideal $I \subseteq k[X_1, X_2, \ldots, X_n]$ have a finite basis and how does one find one?
(2) Given an ideal $I \subseteq k[X_1, X_2, \ldots, X_n]$ and a polynomial f, when is $f \in I$?
(3) Given polynomials $f_1, f_2, \ldots, f_k \in k[X_1, X_2, \ldots, X_n]$, what are the solutions in $k^n$ to the system of equations

$$\begin{aligned} f_1(X_1, X_2, \ldots, X_n) &= 0 \\ \vdots \qquad\qquad & \quad \vdots \\ f_m(X_1, X_2, \ldots, X_n) &= 0? \end{aligned} \tag{10.54}$$

(4) How does one find an implicit representation for a set defined parametrically by

$$\begin{aligned} x_1 &= f_1(t_1, t_2, \ldots, t_n) \\ \vdots \qquad & \quad \vdots \\ x_m &= f_m(t_1, t_2, \ldots, t_n), \end{aligned} \tag{10.55}$$

where the $f_i$ are either polynomials or rational functions?

Our brief introduction to the subject will hopefully motivate the reader to go and learn more about it. Some good references for Gröbner bases and their many applications are [CoLO97], [AdaL94], and [CoLO98].

Gröbner bases were developed to answer questions such as the four listed above in the case of **multi**variable polynomials or ideals where things get more complicated than in the one-variable case. The main reason that answers are either easy or at least well-understood in the 1-variable case is that one has a nice division algorithm. Specif-

**Inputs:** $f(X), g(X) \in k[X]$, $g(X) \neq 0$.
**Outputs:** $a(X), r(X) \in k[X]$ satisfying $f(X) = a(X)\,g(X) + r(X)$ with either $r(X) = 0$ or $\deg r < \deg g$

```
a := 0;  r := f;
while  (r ≠ 0) and (lt(g)  divides lt(r))  do        { lt(p) returns the leading term of p }
    begin
        a := a + lt(r)/lt(g);
        r := r  − (lt(r)/lt(g)) g;
    end;
```

**Algorithm 10.10.1.** The one-variable polynomial division algorithm.

ically, it is known (Theorem B.8.7) that, given polynomials f(X) and g(X), there are unique polynomials a(X) and r(X) with deg r < deg g, so that

$$f(X) = a(X)g(X) + r(X). \tag{10.56}$$

More to the point, there is a simple algorithm for finding a(X) and r(X). Algorithm 10.10.1 is one variant of this one-variable division algorithm. We show it here to help motivate the generalization to the multivariable case that we shall describe shortly. As simple as it is, this division algorithm has many applications. For example, it is used in the Euclidean algorithm that computes the greatest common divisor of a finite set of polynomials. (This algorithm proceeds just like the one for computing the greatest common divisor of a finite set of integers. Structurally, the polynomial ring k[X] looks very much like the integers **Z**, and many of the constructions that one can use for **Z** carry over to k[X].) Algorithm 10.10.1 also leads to simple algorithms that answer questions (1) and (2), because it is the key ingredient to proving that k[X] is a PID (Theorem B.8.11), that is, every ideal $I \subseteq k[X]$ has the form $I = \langle g(X) \rangle$ for some $g \in k[X]$. This obviously shows that every ideal I has a finite basis. A polynomial $f \in k[X]$ belongs to I if and only if the *remainder* r(X) in equation (10.56) is zero.

When we try to answer questions (1) and (2) in the multivariable case, the first problem we encounter is that the ring $k[X_1, X_2, \ldots, X_n]$ is **not** a PID if $n > 1$. Therefore, we shall want to reduce a polynomial by long division with respect to a set of $k > 1$ polynomials (corresponding to the basis of an ideal), not just one. In k[X], this task reduces to the case $k = 1$ because of the PID property. On the other hand, with polynomials of more than one variable it is in general possible to perform long division reductions in different ways which lead to different remainders. Without unique remainders we would lose some good algorithms. Fortunately, we can get uniqueness if we choose the basis carefully (and define a good division algorithm). This is where Gröbner bases come in. The fact that some bases are better than others is hardly new. For example, in the case of vector spaces, orthonormal bases are often particularly desirable.

Next, consider questions (3) and (4). To see the possible role of polynomial division and bases in answers to those questions, we look at the special case of linear equations and parameterizations. The standard solution to the problem in question (3) when we have linear equations is to turn it into a matrix problem and apply the Gauss-Jordan elimination method and standard row reductions. This approach can be thought of as a method for finding certain bases.

**10.10.1. Example.** Consider two linear equations

$$\begin{aligned} f(X,Y,Z) &= X - 2Y + Z = 0 \\ g(X,Y,Z) &= 2X - 3Y + 4Z = 0. \end{aligned} \tag{10.57}$$

Applying the standard row reduction method to the matrix which represents the system of equations (10.57) leads to the row echelon form of that matrix:

$$\begin{pmatrix} 1 & -2 & 1 \\ 2 & -3 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 2 \end{pmatrix}$$

It follows that solving (10.57) is equivalent to solving

$$\begin{aligned} X + 5Z &= 0 \\ Y + 2Z &= 0. \end{aligned} \tag{10.58}$$

Equations (10.58) are easily solved for X and Y in terms of Z and turned into a parametric solution of the form

$$\begin{aligned} X &= -5t \\ Y &= -2t \\ Z &= t. \end{aligned}$$

A similar approach works for question (4) in the linear case.

**10.10.2. Example.** Consider the parameterization

$$\begin{aligned} x_1 &= t_1 + \phantom{2}t_2 - 7 \\ x_2 &= 2t_1 - t_2 + 1 \\ x_3 &= \phantom{2t_1 - {}} 3t_2 - 5, \end{aligned} \tag{10.59}$$

which we rewrite as

$$\begin{aligned} t_1 + t_2 - x_1 \phantom{{}- x_2 - x_3} - 7 &= 0 \\ 2t_1 - t_2 \phantom{{}- x_3} - x_2 \phantom{{}- x_3} + 1 &= 0 \\ 3t_2 \phantom{{}- x_1 - x_2} - x_3 - 5 &= 0. \end{aligned} \tag{10.60}$$

Again, the matrix associated to the system in (10.60) can brought into standard form using row reductions:

$$\begin{pmatrix} 1 & 1 & -1 & 0 & 0 & -7 \\ 2 & -1 & 0 & -1 & 0 & 1 \\ 0 & 3 & 0 & 0 & -1 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{6} & -\frac{2}{3} \\ 0 & 1 & 0 & 0 & \frac{1}{3} & \frac{5}{3} \\ 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} & 5 \end{pmatrix}$$

The last row leads to the equation

$$x_1 - \frac{1}{2}x_2 - \frac{1}{2}x_3 + 5 = 0,$$

which is an implicit representation of the plane defined parametrically by equations (10.59).

Let us see how we can interpret what we did in these examples in polynomial terms. For example, in Example 10.10.1 the row reductions in the matrices correspond to showing that

$$V(f,g) = V(f,g_1) = V(f_1,g_1),$$

where $g_1 = g - 2f$ and $f_1 = f + 2g_1$. Replacing g by $g_1$ will be called a *reduction*. It corresponds to eliminating the X term in g. How this is done is via a long division of g by f:

$$\begin{array}{r} 2 \\ X - 2Y + Z \overline{\big) 2X - 3Y + 4Z} \\ \underline{2X - 4Y + 2Z} \\ Y + 2Z \end{array}$$

If there are more equations, then more divisions may be needed and by more than one polynomial. In analyzing what is going on here, there are two points to observe as we prepare to generalize this process:

(1) We **ordered** the terms of the linear polynomials – we decided to eliminate X first in Example 10.10.1 and we listed the $x_i$ after the $t_i$ in Example 10.10.2.
(2) We subtract multiples of a polynomial to eliminate the current "leading" terms from remaining polynomials.

In the nonlinear case for arbitrary polynomials in $k[X_1,X_2, \ldots ,X_n]$ we shall try to use similar steps, but things get more complicated. We address question (1) first. If an ideal I is generated by polynomials $g_1, g_2, \ldots , g_m$, then to determine whether or not the polynomial f belongs to I reduces to finding polynomials $a_1, a_2, \ldots , a_m$, so that

$$f = a_1g_1 + a_2g_2 + \ldots + a_mg_m. \tag{10.61}$$

Finding such polynomials $a_i$ or showing that they do not exist is not as easy as in the linear case, especially, since we cannot assume anything about their degree even if we

know something about the degrees of f and the $g_i$. For example, if the $g_i$ all have degree 2 and f has degree 4, we cannot assume that the $a_i$ will have degree at most 2. Nevertheless, the approach will be to try to reduce f to 0 by successively subtracting appropriate multiples of $g_1$ from f, then multiples of $g_2$, etc. We will need some criteria that will tell us that as we change f we are making progress. To do this we define a notion for when a term of a polynomial is more complicated than another one. We need an ordering of monomials. One such well-known ordering is the following:

**Definition.** The *lexicographic order* or *lex order* < of the monomials in indeterminates $X_1, X_2, \ldots, X_n$ over a ring R is defined as follows: Given monomials

$$u = cX_1^{e_1}X_2^{e_2}\ldots X_n^{e_n} \quad \text{and} \quad v = dX_1^{f_1}X_2^{f_2}\ldots X_n^{f_n},$$

then u < v if the **left**-most nonzero entry in the vector $\mathbf{d} = (f_1, f_2, \ldots, f_n) - (e_1, e_2, \ldots, e_n)$ in $\mathbf{Z}^n$ is positive. The *reverse lexicographic order* defines u < v if the **right**-most nonzero entry in the vector **d** is negative.

For example, if $X = X_1$ and $Y = X_2$, then

$$1 < Y < Y^2 < \ldots < Y^k < \ldots < X < XY < XY^2 < \ldots < X^2 < X^2Y < \ldots$$

with respect to the lexicographic order. Note that the lexicographic order is determined by the order in which the indeterminates $X_i$ are listed. In essence, one has chosen a fixed ordering of the variables, namely, $X_n < X_{n-1} < \ldots < X_1$.

The lexicographic order is only one of many orderings. Other orderings actually work better with certain problems. Here are two other standard ones; however, the reader should be aware of the fact that the terminology varies between authors.

**Definition.** The *degree lexicographic order* or *deglex order* < of the monomials in indeterminates $X_1, X_2, \ldots, X_n$ over a ring R is defined by saying that any monomial of total degree d is less than any monomial of degree e if d < e and monomials of equal total degree are ordered using the lexicographic order.

For example, if $X = X_1$ and $Y = X_2$, then

$$1 < Y < X < Y^2 < XY < X^2 < \ldots.$$

**Definition.** The *degree reverse lexicographic order* or *degrevlex order* < of the monomials in indeterminates $X_1, X_2, \ldots, X_n$ over a ring R is defined by saying that any monomial of total degree d is less than any monomial of degree e if d < e and monomials of equal total degree are ordered using the reverse lexicographic order.

In the case of two variables, the degrevlex and deglex order are the same (Exercise 10.10.2). On the other hand, when $X_3 < X_2 < X_1$, then

$$X_1X_2^3 < X_1^2X_2X_3$$

for deglex order, but

$$X_1^2X_2X_3 < X_1X_2^3$$

for degrevlex order.

We would like to isolate the essential properties that an ordering should possess to be useful. Since ordering monomials

$$cX_1^{e_1}x_2^{e_2}\ldots X_n^{e_n}$$

will ignore the constant c and will be equivalent to ordering their exponent tuples

$$\mathbf{e} = (e_1, e_2, \ldots, e_n)$$

in $\mathbf{N}^n$, the definition of an ordering often deals simply with orderings of elements of $\mathbf{N}^n$ rather than monomials. We shall state both definitions for comparison purposes but concentrate on $\mathbf{N}^n$.

**Definition.** A *monomial ordering* of $\mathbf{N}^n$ is a total ordering $<$ on $\mathbf{N}^n$ that satisfies

(1) $\mathbf{0} \le \mathbf{e}$ for all $\mathbf{e} \in \mathbf{N}^n$, and
(2) if whenever $\mathbf{e}, \mathbf{f} \in \mathbf{N}^n$ and $\mathbf{e} < \mathbf{f}$, then $\mathbf{e}+\mathbf{g} < \mathbf{f}+\mathbf{g}$ for all $\mathbf{g} \in \mathbf{N}^n$.

**Definition.** A *monomial ordering* of $k[X_1,X_2,\ldots,X_n]$ is a total ordering $<$ on the monomials of $k[X_1,X_2,\ldots,X_n]$ satisfying

(1) $c < X_i$ for all $c \in k$ and all i.
(2) For all monomials u, v, $w \in k[X_1,X_2,\ldots,X_n]$, $u < v$ implies that $uw < vw$.

**10.10.3. Theorem.** Every monomial ordering of $\mathbf{N}^n$ is a well-ordering.

**Proof.** Let $<$ be monomial ordering. We need to show that every subset of $\mathbf{N}^n$ has a smallest element, or, equivalently, that every decreasing sequence

$$\ldots < \mathbf{e}_{i+1} < \mathbf{e}_i < \ldots < \mathbf{e}_2 < \mathbf{e}_1, \quad \mathbf{e}_i = (e_{i1}, e_{i2}, \ldots, e_{in}) \in \mathbf{N}^n, \tag{10.62}$$

must terminate after a finite number of steps. The theorem will be proved by induction on n.

If $n = 1$, then it is easy to show that $<$ is just the standard less than relation on the nonnegative integers. Assume that $n \ge 2$ and that the theorem is true for $n - 1$. Suppose that we have a decreasing sequence of elements $\mathbf{e}_i$ as shown in (10.62).

**Claim 1.** The theorem is true if there is a j so that all the jth entries in the $\mathbf{e}_i$ are constant.

Without loss of generality we may assume that $j = n$. Suppose that $e_{in} = c$ for all i. Because

$$(f_{i1}, f_{i2}, \ldots, f_{in-1}, c) < (g_{i1}, g_{i2}, \ldots, g_{in-1}, c) \quad \text{if and only if}$$
$$(f_{i1}, f_{i2}, \ldots, f_{in-1}, 0) < (g_{i1}, g_{i2}, \ldots, g_{in-1}, 0),$$

we may also assume that c = 0. Define an ordering < of $\mathbf{N}^{n-1}$ by:

$$(f_{i1}, f_{i2}, \ldots, f_{in-1}) < (g_{i1}, g_{i2}, \ldots, g_{in-1}) \quad \text{if and only if}$$
$$(f_{i1}, f_{i2}, \ldots, f_{in-1}, 0) < (g_{i1}, g_{i2}, \ldots, g_{in-1}, 0). \tag{10.63}$$

It is easy to check that this defines a monomial ordering on $\mathbf{N}^{n-1}$. Our inductive hypothesis applied to the sequence $\mathbf{e}_i' = (e_{i1}, e_{i2}, \ldots, e_{in-1})$ in now proves Claim 1.

**Claim 2.** If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbf{N}^{n-1}$ satisfies $x_j < y_j$ for all j, then $\mathbf{x} < \mathbf{y}$.

Let $d_j = y_j - x_j$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$. Then $\mathbf{0} < \mathbf{d}$ implies that $\mathbf{x} = \mathbf{0} + \mathbf{x} < \mathbf{d} + \mathbf{x} = \mathbf{y}$ and Claim 2 is proved.

Now consider the sequence in (10.62) again and assume that it is an infinite sequence. If any of the sequences of nonnegative integers $e_{ij}$, j = 1,2, . . . , is bounded for some fixed i, then we can choose a subsequence of the $\mathbf{e}_i$ so that the jth entries in that subsequence are constant. By Claim 1 we would be done. Assume therefore that the sequences $e_{ij}$, j = 1,2, . . . , are unbounded for all i. Choose a subsequence of the $\mathbf{e}_i$ so that for that subsequence the first components form a strictly increasing sequence of integers going to infinity. Let $\mathbf{f}_i$ be this new decreasing sequence of elements of $\mathbf{N}^n$. Now look at the second components of all the $\mathbf{f}_i$ and choose a subsequence so that the second components form a strictly increasing sequence of integers going to infinity. Continue in this way until we finally end up with a subsequence $\mathbf{g}_i = (g_{i1}, g_{i2}, \ldots, g_{in})$ of the original sequence $\mathbf{e}_i$ with the property that $g_{i+1,j} > g_{ij}$ for all i and j. By Claim 2 we would have $\mathbf{g}_{i+1} > \mathbf{g}_i$, which is impossible since the sequence $\mathbf{e}_i$ was decreasing. This contradiction proves the theorem.

Theorem 10.10.3 is important because we need monomial orderings to have the well-ordering property in addition to the total order to ensure that various algorithms we shall define will terminate.

**10.10.4. Proposition.** The lex, deglex, and degrevlex orders of $k[X_1, X_2, \ldots, X_n]$ are monomial orderings.

**Proof.** Obvious.

We need some more terminology before we can state the multivariable long division theorem.

**Definition.** Fix a monomial order < on $k[X_1, X_2, \ldots, X_n]$ and let $f \in k[X_1, X_2, \ldots, X_n]$. The largest monomial in f with respect to this order is called the *leading term* of f and is denoted by lt(f). Its coefficient is called the *leading coefficient* and is denoted by lc(f). The *leading power product*, denoted by lpp(f), is defined by the equation lt(f) = lc(f)lpp(f) and is the largest power product appearing in f.

For example, assuming deglex order and Y < X, if

$$f(X, Y) = 7X^2Y + XY - Y,$$

then

$$\mathrm{lt}(f) = 7X^2Y, \quad \mathrm{lc}(f) = 7, \quad \text{and} \quad \mathrm{lpp}(f) = X^2Y.$$

In the future, whenever we talk about leading terms, etc., we shall always assume a monomial order has been chosen even if we do not say so explicitly.

**10.10.5. Theorem.** (The Division Algorithm for $k[X_1,X_2,\ldots,X_n]$) Fix a monomial order $<$ on $k[X_1,X_2,\ldots,X_n]$ and let $P = \{p_1,p_2,\ldots,p_m\}$ be a fixed set of m nonzero polynomials in $k[X_1,X_2,\ldots,X_n]$. Then every $f \in k[X_1,X_2,\ldots,X_n]$ can be expressed as

$$f = a_1p_1 + a_2p_2 + \ldots + a_mp_m + r,$$

where $a_i$, $r \in k[X_1,X_2,\ldots,X_n]$ , and either r is zero or none of the monomials appearing in r is divisible by any of the $\mathrm{lt}(p_i)$. The polynomial r will be called a *remainder* of f by division with respect the sequence of polynomials P. Furthermore,

$$\mathrm{lpp}(f) = \max\{\mathrm{lpp}(a_1)\,\mathrm{lpp}(p_1), \mathrm{lpp}(a_2)\,\mathrm{lpp}(p_2), \ldots, \mathrm{lpp}(a_m)\,\mathrm{lpp}(p_m), \mathrm{lp}(r)\}.$$

**Proof.** Algorithm 10.10.2 is an algorithm that finds the $a_i$ and r. Therefore, the proof of the division theorem boils down to showing that that algorithm does what it claims. The reader should compare this algorithm with Algorithm 10.10.1 for one variable. Two key observations for a proof of Algorithm 10.10.2 are:

(1) $f = a_1p_1 + a_2p_2 + \cdots + a_mp_m + g + r$ at each stage.
(2) The leading term of g in (1) decreases with respect to the ordering so that the algorithm terminates.

See [CoLO97] or [AdaL94].

See Exercise 10.10.4 for a simple variant of a multivariable division algorithm.

**Definition.** Let f and g be two polynomials. We say that f is *simpler* than g if $\mathrm{lt}(f) < \mathrm{lt}(g)$.

For example, if

$$f = XY^2 + X^3 \quad \text{and} \quad g = X^3Y + Y^3,$$

then f is simpler than g. Algorithm 10.10.2 shows how to simplify polynomials with respect to any set of polynomials. We introduce some other common terminology.

**Definition.** Fix a monomial order. Let P be a set of polynomials and f an arbitrary polynomial. Assume that some monomial term u of f is divisible by the leading term of a polynomial p in P, that is,

**Inputs:** $f, p_1, p_2, \ldots, p_m \in k[X_1, X_2, \ldots, X_n]$ , $p_i \neq 0$
**Outputs:** $a_1, a_2, \ldots, a_m, r \in k[X_1, X_2, \ldots, X_n]$ satisfying

$$f = a_1 p_1 + \ldots + a_m p_m + r,$$

where either r is zero or none of the monomials appearing in r is divisible by any of the $lt(p_i)$

```
integer i;
boolean noDivision;

a1 := a2 := ...:= am := r := 0;    g := f;
while g ≠ 0 do
    begin
        i := 1;   noDivision := true;
        while (i ≤ m) and noDivision do
            if lt(pi) divides lt(g)
                then
                    begin
                        ai := ai + lt(g)/lt(pi);
                        g := g − (lt(g)/lt(pi))*pi;
                        noDivision := false;
                    end
                else i := i + 1;

        if noDivision then
            begin
                r := r + lt(g);
                g := g − lt(g);
            end
    end;
```

**Algorithm 10.10.2.** The multivariable polynomial division algorithm.

$$u = h\, lt(p)$$

for some monomial h. Let $g = f - hp$. We shall say that the polynomial g is obtained from f via an *elementary P-reduction* and write $f \Rightarrow g$. If no such polynomial p exists, then we shall say that f is *P-irreducible*. If there exists a sequence of elementary P-reductions $f = f_0 \Rightarrow f_1 \Rightarrow f_2 \Rightarrow \ldots \Rightarrow f_m = g$, then we say that f *P-reduces* to g and write $f \xrightarrow{P} g$. If f P-reduces to g and g is P-irreducible, we shall call g a *P-normal form* for f and denote such g by NF(f,P).

The term "P-normal form" for a polynomial is **not quite** just another name for a remainder in Algorithm 10.10.2. The reason is that an elementary reduction of f does not require that the term u in f is a leading term. In practical algorithms, however, u always **will** be chosen to be a leading term because of Algorithm 10.10.2.

**10.10.6. Example.** Choose the lexicographic order for k[X,Y] with X > Y and let

$$P = \{\, p_1 = 3X^2Y + X,\ p_2 = X^2Y^2 - X,\ p_3 = Y,\ p_4 = 4X^2 + X \,\}$$

and

$$f = 12X^3Y - X^2Y^2 + 9XY^2.$$

The first term of each of the polynomials $p_i$ is the leading term. Define polynomials $g_i$ and $h_i$ by

$$\begin{aligned}
g_1 &= f - 4Xp_1 = -X^2Y^2 - 4X^2 + 9XY^2, & h_1 &= f - 12X^3p_3 = -X^2Y^2 + 9XY^2\\
g_2 &= g_1 + p_2 = -4X^2 + 9XY^2 - X, & h_2 &= h_1 + p_2 = 9XY^2 - X,\\
g_3 &= g_2 - 9XY\, p_3 = -4X^2 - X, & h_3 &= h_2 - 9XY\, p_3 = -X,\\
g_4 &= g_3 + p_4 = 0.
\end{aligned}$$

This shows that

$$f \Rightarrow g_1 \Rightarrow g_2 \Rightarrow g_3 \Rightarrow g_4 \quad \text{and} \quad f \Rightarrow h_1 \Rightarrow h_2 \Rightarrow h_3.$$

Thus, f P-reduces to $g_4$ and $h_3$. Since both $g_4$ and $h_3$ are P-irreducible, they are both P-normal forms for f.

Note that Example 10.10.6 shows that P-normal forms (or remainders in Algorithm 10.10.2) are **not** unique in general. Now, an elementary P-reduction introduces potentially new terms to the original polynomial. It might seem initially that we could have an infinite chain of elementary P-reductions. This is not possible however. See [AdaL94]. (The termination part of the proof for Algorithm 10.10.2 does not apply directly because elementary P-reductions of f do not necessarily pick a leading term of f. One needs a simple modification of that proof.)

**10.10.7. Proposition.** Let P be a finite set of polynomials and f an arbitrary polynomial. If the P-normal form g for f is unique, then g = 0 if and only if f belongs to the ideal <P>.

**Proof.** Let $P = \{p_1, p_2, \ldots, p_m\}$. Now, if $f \xrightarrow{P} g$, then

$$f = a_1p_1 + a_2p_2 + \ldots + a_mp_m + g$$

for some polynomials $a_i$. Therefore, if g = 0 then $f \in \langle P \rangle$. Conversely, let $f \in \langle P \rangle$. It is not true in general that every P-normal form of f is 0. See Exercise 10.10.6. On the other hand, at least one will be because if

$$f = a_1p_1 + a_2p_2 + \ldots + a_mp_m$$

for some polynomials $a_i$ (each of which is a sum of monomials), then f is sum of products $b_jq_j$, where $b_j$ is a monomial and $q_j \in P$, and this provides a sequence of elementary P-reductions that lead to 0. It follows that if the P-normal form is unique, then all P-normal forms will be 0.

Proposition 10.10.7 motivates us to find bases P for ideals so that they induce unique P-normal forms on polynomials. Such bases will be called Gröbner bases. There are many possible equivalent definitions for these. We choose one based on leading term properties rather than the unique P-normal form property because it is the former that play the central role in all the proofs and practical algorithms.

**Definition.** Fix a monomial order and let I be a nonzero ideal in $k[X_1,X_2, \ldots ,X_n]$. A finite set $P = \{p_1,p_2, \ldots ,p_m\}$ of nonzero polynomials in I is called a *Gröbner basis* or *standard basis* for I if and only if for all nonzero $f \in I$, there is some j, $1 \le j \le n$, so that $lt(p_j)$ divides $lt(f)$.

Note that if we have a Gröbner basis P for an ideal I, then no nonzero polynomial in I is P-irreducible.

**Definition.** Let S be an arbitrary nonempty subset of $k[X_1,X_2, \ldots ,X_n]$. Define the set of leading terms of S, lt(S), by

$$lt(S) = \{lt(f) \mid f \in S\}.$$

**10.10.8. Theorem.** Fix a monomial order and let I be a nonzero ideal in $k[X_1,X_2, \ldots ,X_n]$. The following statements are equivalent for a finite set $P = (p_1,p_2, \ldots ,p_m)$ of nonzero polynomials in I:

(1) P is a Gröbner basis for I.
(2) $f \in I$ if and only if $f \xrightarrow{P} 0$.
(3) $f \in I$ if and only if

$$f = \sum_{i=1}^{m} a_ip_i \quad \text{and} \quad lpp(f) = \max_{1\le i\le n} \{lpp(a_i)lpp(p_i)\},$$

for some polynomials $a_i$.
(4) $\langle lt(P)\rangle = \langle lt(I)\rangle$.
(5) Every polynomial $f \in k[X_1,X_2, \ldots ,X_n]$ has a unique P-normal form.

**Proof.** We shall only prove that (1)–(4) are equivalent. For a proof showing that (5) is equivalent to (2) see [AdaL94].

(1) $\Rightarrow$ (2): Let $f \in I$. Let $r = NF(f,P)$. Clearly, $r \in I$. But r must be 0, otherwise one would be able to reduce it further by the definition of a Gröbner basis. The converse is also immediate.

(2) ⇒ (3): This is a straightforward consequence of Theorem 10.10.5.

(3) ⇒ (4): Obviously, $<lt(P)> \subseteq <lt(I)>$. We need to show the reverse inclusion. Let $g \in <lt(I)>$. Then g is a linear combination of leading terms of polynomials in I. But if $f \in I$, then (3) implies that

$$lt(f) = \sum_i lt(a_i)lt(p_i),$$

where the sum is taken over all i with the property that

$$lpp(f) = lpp(a_i)lpp(p_i).$$

This shows that g is a linear combination of $lt(p_i)$, that is, $g \in <lt(P)>$.

(4) ⇒ (1): Let $f \in I$. By (4), lt(f) can be expressed in the form

$$lt(f) = \sum_{i=1}^{m} b_i lt(p_i),$$

where $b_i \in k[X_1,X_2, \ldots ,X_n]$. Consideration of the right-hand side of this equation after expanding the polynomials $b_i$ into their monomial parts shows that lpp(f) must be divisible by $lpp(p_j)$ for some j, which is what we had to prove.

**10.10.9. Corollary.** If $P = \{p_1,p_2, \ldots ,p_m\}$ is a Gröbner basis for an ideal I, then $I = <p_1,p_2, \ldots ,p_m>$.

**Proof.** Since each $p_i$ belongs to I, we clearly have $<p_1,p_2, \ldots ,p_m> \subseteq I$. On the other hand, if $f \in I$, then Theorem 10.10.8(3) implies that $f \in <p_1,p_2, \ldots ,p_m>$.

We have listed some properties of Gröbner bases, but do they exist?

**10.10.10. Theorem.** Every nonzero ideal in $k[X_1,X_2, \ldots ,X_n]$ has a Gröbner basis.

**Proof.** See [CoLO97] or [AdaL94]. One needs the Hilbert basis theorem for this.

The next question is how one can find a Gröbner basis. Consider an ideal $I = <p_1,p_2, \ldots ,p_m>$ and let $P = \{p_1,p_2, \ldots ,p_m\}$. What might cause P not to be a Gröbner basis for I? First of all, although every element f in I is a linear combination of the $p_i$, the leading terms might cancel out leaving f with a leading term that is smaller than all of the leading terms of the $p_i$.

**10.10.11. Example.** Using the deglex order of k[X,Y] with X > Y, let

$$I = < p_1 = X^2Y + X, p_2 = XY^2 + 1 > \quad \text{and} \quad f = Yp_1 - Xp_2 = XY - X \in I.$$

The existence of such a polynomial f in I shows that the set $P = \{p_1,p_2\}$ does not form a Gröbner basis.

The next observation strikes more at the heart of the problem with picking any old basis for an ideal. The problem is that one usually has many choices when reduc-

ing polynomials to get to a reduced one. For example, there are two choices when reducing the polynomial

$$g = X^2Y^2 - X$$

with respect to the set of polynomials P in Example 10.10.11, that is, g has the two P-normal forms

$$g_1 = g - yp_1 = -XY - X \quad \text{and} \quad g_2 = g - xp_2 = -2X.$$

That one gets different sequences of polynomials on the way to a reduced form might not be bad by itself, but what turns out to be bad is that differences such as

$$g_2 - g_1 = XY - X \tag{10.64}$$

are irreducible. Trying to fix this problem is what leads to an algorithm for finding Gröbner bases.

**Definition.** Let f and g be nonzero polynomials in $k[X_1,X_2,\ldots,X_n]$. Let L = lcm(lpp(f),lpp(g)). The *S-polynomial* of f and g, denoted by S(f,g), is defined by

$$S(f,g) = \frac{L}{lt(f)}f - \frac{L}{lt(g)}g.$$

For example, if $p_1$ and $p_2$ are the polynomials in Example 10.10.11, then

$$S(p_1,p_2) = Yp_1 - Xp_2 = XY - X.$$

This is in fact the same polynomial as the one in equation (10.64) that measured differences in outcomes of reductions.

**10.10.12. Theorem.** (Buchberger) Let P be a finite set of polynomials in $k[X_1,X_2,\ldots,X_n]$. Then the following are equivalent:

(1) P is a Gröbner basis for <P>.
(2) For all f and g in P we have NF(S(f,g),P) = 0.

**Proof.** See [CoLO97] or [AdaL94].

**10.10.13. Theorem.** (Buchberger Gröbner Basis Algorithm) Fix a monomial order and let I be a nonzero ideal in $k[X_1,X_2,\ldots,X_n]$. If $P = \{p_1,p_2,\ldots,p_m\}$ is any basis for I, that is, $I = \langle p_1,p_2,\ldots,p_m\rangle$, then Algorithm 10.10.3 computes a Gröbner basis G for I.

**Proof.** See [CoLO97] or [AdaL94]. Algorithm 10.10.3 does not produce a unique Gröbner basis because different choices could be made as it proceeds.

**10.10.14. Example.** To use Algorithm 10.10.3 to find a Gröbner basis for the ideal $I = \langle p_1 = X^2Y + X, p_2 = XY^2 + 1\rangle$ using the deglex order on k[X,Y] with X > Y.

```
Input:      A nonempty finite set P of polynomials.
Output:     A Gröbner basis G for the ideal <P>.

polynomialSet      G;
polynomialPairSet  B;
polynomial         p, q, g;

G := P;   B := { {p,q} | p, q ∈ G and p ≠ q } ;
while B ≠ ϕ do
    begin
        {p,q} := AnyElementOf (B);
        B := B − {{p,q}};
        g := NF(S(p,q),G);
          if g ≠ 0 then
            begin
                B := B ∪{ {g,g′ } | g′ ∈ G };
                G := G ∪{ g };
            end
    end;
```

**Algorithm 10.10.3.** The Buchberger Gröbner basis algorithm.

**Solution.** We show the steps in the algorithm.

At start of first time through loop: $G = \{p_1,p_2\}$ and $B = \{\{p_1,p_2\}\}$
In first loop: $\{p,q\} = \{p_1,p_2\}$, B is set to empty,
$S(p,q) = XY - X$, and $g = NF(S(p,q)) = XY - X$.
We add to B and G

At start of second time through loop: $G = \{p_1,p_2,XY - X\}$ and
$B = \{\{XY - X,p_1\}, \{XY - X,p_2\}\}$
In second loop: $(p,q) = \{XY - X,p_1\}$, B is set to $\{\{XY - X,p_2\}\}$
$S(p,q) = -XY - X$, and $g = NF(S(p,q)) = -2X$.
We add to B and G

At start of third time through loop: $G = \{p_1,p_2,XY - X,-2X\}$ and
$B = \{\{-2X,p_1\},\{-2X,p_1\},\{-2X,XY - X\},\{XY - X,p_2\}\}$
In third loop: $\{p,q\} = \{-2X,p_1\}$, B is set to $\{\{-2X,p_1\},\{-2X,XY - X\},\{XY - X,p_2\}\}$,
$S(p,q) = X$, and $g = NF(S(p,q)) = 0$.

We now go through the loop three more times, removing one element of B each time, and each time the reduced element g is again 0.

This shows that the algorithm produces the basis $G = \{p_1, p_2, XY - X, -2X\}$, which is easily shown to satisfy the definition of a Gröbner basis. Of course, since there were choices made along the way (we could have chosen different elements of B), we should not expect to end up with a unique basis.

Because finding a Gröbner basis can be very time consuming, it is important that one proceeds in as efficient manner as possible. The basic Buchberger algorithm is not very efficient. For one thing, the set B may get large. We shall briefly mention two ways to improve the algorithm. The first simplification of Gröbner bases rests on the following lemma.

**10.10.15. Lemma.** If G is a Gröbner basis for an ideal I and if $p \in G$ is a polynomial with the property that $\mathrm{lt}(p) \in \langle \mathrm{lt}(G - \{p\}) \rangle$, then $G - \{p\}$ is also a Gröbner basis for I.

**Proof.** Easy.

**Definition.** A Gröbner basis $G = \{g_1, g_2, \ldots, g_m\}$ is called *minimal* if

(1) $\mathrm{lc}(g_i) = 1$ for all i.
(2) For all i and j, $i \neq j$, $\mathrm{lpp}(g_i)$ does not divide $\mathrm{lpp}(g_j)$.

For example, given the Gröbner basis

$$G = \{X^2Y + X, XY^2 + 1, XY - X, -2X\}$$

that we got in Example 10.10.14, we can normalize the fourth element and apply Lemma 10.10.15 to the first three elements of G to show that $\{X\}$ is also a Gröbner basis that is in fact minimal.

Minimal Gröbner bases get rid of some extra elements but may not be unique. For example, the ideals $\langle X^2 + aXY, XY \rangle$ are minimal Gröbner bases for the ideal $\langle X^2, XY \rangle$ for any constant $a \in k$.

**Definition.** A Gröbner basis $G = \{g_1, g_2, \ldots, g_m\}$ is said to be *reduced* if

(1) $\mathrm{lc}(g_i) = 1$ for all i.
(2) For all i, $g_i$ is reduced with respect to $G - \{g_i\}$. Equivalently, no nonzero term in $g_i$ is divisible by any $\mathrm{lpp}(g_j)$ for $j \neq i$.

**10.10.16. Theorem.** If we fix a monomial order, then every nonzero ideal I in $k[X_1, X_2, \ldots, X_n]$ has a **unique** reduced Gröbner basis with respect to that order.

**Proof.** See [CoLO97] or [AdaL94].

One immediate application of the uniqueness in Theorem 10.10.16 is the following:

**A test for when ideals are equal:** Simply compute the reduced Gröbner basis for both and check if they are the same.

Another application using Gröbner bases is:

**A test for ideal membership:** To see if a polynomial f belongs to an ideal, compute a Gröbner basis G for the ideal and check if the G-normal form for f is zero.

Here are some more uses of Gröbner bases:

**10.10.17. Example.** To find the solutions **S** to the equations

$$\begin{aligned} x^2 + y^2 - z^2 &= 1 \\ x^2 - z^2 &= y \\ x^2 &= y. \end{aligned}$$

**Solution.** If

$$p_1 = x^2 + y^2 - z^2 - 1, \quad p_2 = x^2 - y - z^2, \quad \text{and} \quad p_3 = x^2 - y,$$

then by Lemma 10.8.2, **S** = $V(p_1,p_2,p_3) = V(I)$, where $I = \langle p_1,p_2,p_3\rangle$. Using our algorithms one can show that $G = \{g_1,g_2,g_3\}$ is a reduced Gröbner basis for I, where

$$\begin{aligned} g_1 &= x^2 - y \\ g_2 &= y^2 + y - 1 \\ g_3 &= z^2. \end{aligned}$$

Since $I = \langle g_1,g_2,g_3\rangle$, the interesting thing about this is that $g_2$ and $g_3$ are polynomials of only **one** variable. Solving for their zeros we get

$$\begin{aligned} z &= 0 \\ y &= \frac{-1 \pm \sqrt{5}}{2}. \end{aligned}$$

Substituting these solutions into the equation $g_1 = 0$ and solving for x gives

$$x = \pm\sqrt{\frac{-1 \pm \sqrt{5}}{2}}.$$

These values give us all the answers over the complex numbers. Although our example is a trivial one, we shall see that it is not atypical and that Gröbner bases can be used to eliminate variables in equations, solve the simpler equations, and then extend the partial solutions to a solution of the original equations.

**10.10.18. Example.** To find an implicit form for the curve parameterized by

$$\begin{aligned} x &= t^4 \\ y &= t^3 \\ z &= t^2. \end{aligned}$$

**Solution.** Clearly, the curve is just the set

$$V(t^4 - x, t^3 - y, t^2 - z) = V(I),$$

where I is the ideal $<t^4 - x, t^3 - y, t^2 - z>$ in $\mathbf{C}[t,x,y,z]$. One can show that

$$G = \{-t^2 + z, ty - z^2, tz - y, x - z^2, y^2 - z^3\}$$

is a Gröbner basis for I using the lex order of $\mathbf{C}[t,x,y,z]$ with $z < y < x < t$. Note how the last two polynomials in G do not involve the parameter t. It follows that the curve is contained in the variety defined by

$$\begin{aligned} x - z^2 &= 0 \\ y^2 - z^3 &= 0. \end{aligned}$$

Does this variety in $\mathbf{C}^3$ contain any extra points that were not in the curve? It turns out the answer to this question is no. See Theorem 10.11.4 and the comments leading up to it. We were again able to solve our problem with Gröbner bases. The general algorithm for the implicitization problem using this approach is described in the next section.

In this section we have seen how useful Gröbner bases can be. It is therefore important to have the most efficient algorithm for computing them possible. We have indicated a few improvements to the basic Buchberger algorithm. More improvements can be made by a careful analysis of the role that S-polynomials play in the algorithm. We refer the reader to [AdaL94] or [CoLO97] for a discussion of such improvements and actual algorithms that carry them out. Even so, finding Gröbner bases can still be a very slow process for certain ideals. From a theoretical point, it is a very complex problem for the worst cases. One problem is that the degrees of the polynomials in a Gröbner basis can be quite large. It has been shown that even if an ideal can be generated by polynomials whose total degree is bounded by some d, the degrees of the polynomials in a Gröbner bases for it can in certain cases be doubly exponential in d. Initial pessimism in this regard has been mitigated by the fact that things do not turn out so badly in many practical problems. Recall our pointing out that the monomial order we choose and the way we order the individual variables $X_i$ can influence the results. One has found that the degrevlex order seems to work best in most instances in that one seems to get polynomials of smallest total degree. Nevertheless, although Gröbner bases have a great theoretical value and have much going for them, it turns out that methods based on resultants turn out to be more efficient in many practical problems. Sweeping statements that one method is always better than the other are not possible.

The implicitization problem remains a difficult computational problem in general. The solution can end up being a high degree polynomial with a large number of terms. For example, in CAGD a triangular surface patch of degree n has an implicit equation of degree $n^2$ in general. A tensor product patch using polynomials of degree n and m has an implicit equation of degree 2nm in general. See [Sede87]. For another approach to implicitization using the Wu-Ritt method see [Hoff93] or [CoLO97].

## 10.11 Elimination Theory

Resultants and Gröbner bases are two very useful and general tools in algebraic geometry. In the last two sections we saw how they could be applied to the problem of implicitizing parameterized curves and solving equations. The approach basically boiled down to eliminating variables and extending solutions. The examples we gave, however, were just that, examples, and without any theory about the elimination process and extending solutions in a systematic way. We correct that now and, at least briefly, indicate some essential results in that theory. A more thorough discussion can be found in [CoLO97].

**Definition.** Given an ideal $I = \langle f_1, f_2, \ldots, f_n \rangle \subseteq k[X_1, X_2, \ldots, X_n]$, define the mth *elimination ideal* $I_m$ by

$$I_m = I \cap k[X_{m+1}, X_{m+2}, \ldots, X_n].$$

Clearly, $I_m$ consists of polynomials in I from which the variables $X_1, X_2, \ldots, X_m$ have been eliminated. In other words, to eliminate those variables we need to find nonzero polynomials in $I_m$.

**10.11.1. Theorem.** (The Elimination Theorem) If $I \subseteq k[X_1, X_2, \ldots, X_n]$ is an ideal and if G is a Gröbner basis for I with respect to the lex order, where $X_n < X_{n-1} < \cdots < X_1$, then the set

$$G_m = G \cap k[X_{m+1}, X_{m+2}, \ldots, X_n]$$

is a Gröbner basis of the mth elimination ideal $I_m$ for all m, $0 \le m \le n$.

**Proof.** See [CoLO97].

Theorem 10.11.1 tells us something about how to eliminate variables from equation. When it comes to finding solutions to a set of equations, the elimination ideals suggest an approach with the following inductive step: If we have a solution for the equations corresponding to the elimination ideal $I_j$, we extend this solution to a solution for the equations in one more variable corresponding to $I_{j-1}$. Unfortunately, such an extension may not always exist.

**10.11.2. Example.** Consider the equations

$$\begin{aligned} xy &= 1 \\ xz &= 1. \end{aligned}$$

These equations have solution set

$$\mathbf{X} = V(xy-1, xz-1) = V(I), \quad \text{where} \quad I = \langle xy-1, xz-1 \rangle.$$

The Elimination Theorem shows that the elimination ideal $I_1 = \langle y - z \rangle$. All solutions (c,c) to the equation y = z, **except** (0,0), extend to solutions $(1/c, c, c) \in \mathbf{X}$.

The next theorem answers the question about when extensions exist in one case.

**10.11.3. Theorem.** (The Extension Theorem) Assume that k is an algebraically closed field. Let $I = \langle f_1, f_2, \ldots, f_m \rangle \subseteq k[X_1, X_2, \ldots, X_n]$ and let $I_1$ be the first elimination ideal of I. For each $f_i$, $1 \le i \le m$, write $f_i$ in the form

$$f_i = g_i(X_2, \ldots, X_n)X_1^{N_i} + \text{terms in which } X_1 \text{ has degree less than } N_i,$$

where $N_i \ge 0$ and $g_i \in \mathbf{k}[X_2, \ldots, X_n]$ is nonzero. Suppose that we have a partial solution $(a_2, \ldots, a_n) \in V(I_1)$. If $(a_2, \ldots, a_n) \notin V(g_1, g_2, \ldots, g_m)$, then there exists an $a_1 \in k$, such that $(a_1, a_2, \ldots, a_n) \in V(I)$.

**Proof.** See [CoLO97].

The reader should look at Examples 10.10.17 and 10.10.18 again to see how the last two theorems justify the Gröbner basis approach to solving the problems in those examples. A geometric interpretation of Theorem 10.11.3 can be found in [CoLO97].

Next, we address a question raised in the solution for Example 10.10.18 that had to do with the nature of the implicit representation that we obtain for a set **X** from a given parametric one using the Gröbner basis approach. Let **V** be the variety that one obtains. The set **X** is presumably contained in **V** by construction, but since the set **X** might not even be a variety, one should not expect **V** to be the same as **X** in general. On the other hand, the variety **V** could easily be much larger than necessary. Therefore, the implicitization problem should be stated more precisely as asking for the **smallest** variety that contains **X**. The following questions need to be answered:

(1) Is the variety **V** obtained via the Gröbner basis approach this smallest variety?
(2) If $\mathbf{V} \ne \mathbf{X}$, then how does one determine those points of **V** not in **X**?

The first of these questions is answered by the following theorem.

**10.11.4. Theorem.** (The Polynomial Implicitization Theorem) Let k be an infinite field. Consider the parameterization defined in equation (10.55) and define $F: k^n \to k^m$ by

$$F(t_1, t_2, \ldots, t_n) = (f_1(t_1, t_2, \ldots, t_n), f_2(t_1, t_2, \ldots, t_n), \ldots, f_m(t_1, t_2, \ldots, t_n)).$$

Let I be the ideal

$$\langle x_1 - f_1, x_2 - f_2, \ldots, x_m - f_m \rangle \subseteq k[t_1, \ldots, t_n, x_1, \ldots x_m]$$

and let

$$I_n = I \cap k[x_1, x_2, \ldots, x_m]$$

be the nth elimination ideal. Then $V(I_n)$ is the smallest variety in $k^m$ containing $F(k^n)$.

**Proof.** See [CoLO97].

**Input:** A set $\mathbf{X}$ in $k^m$ with parameterization

$$\begin{aligned} x_1 &= f_1(t_1,t_2,\ldots,t_n), \\ x_2 &= f_2(t_1,t_2,\ldots,t_n), \\ &\vdots \\ x_m &= f_m(t_1,t_2,\ldots,t_n), \end{aligned}$$

where $f_i \in k[t_1, \ldots, t_n]$

**Output:** A set of polynomials $g_1, g_2, \ldots, g_s \in k[x_1, \ldots, x_m]$, so that the variety $\mathbf{V} = V(g_1,g_2,\ldots,g_s)$ is the smallest variety in $k^m$ containing $\mathbf{X}$.

Compute a Gröbner basis for the ideal $I = \langle x_1 - f_1, x_2 - f_2, \ldots, x_m - f_m \rangle$ in $k[t_1,\ldots,t_n,x_1,\ldots x_m]$ with respect to the lex order, where

$$x_m < x_{m-1} < \cdots < x_1 < t_n < t_{n-1} < \cdots < t_1 .$$

The elements $g_j$ of the Gröbner basis not involving the variables $t_i$ are the basis of an ideal J and $\mathbf{V} = V(J)$.

**Algorithm 10.11.1.** Implicitization algorithm using Gröbner bases.

Theorems 10.11.1 and 10.11.4 lead to Algorithm 10.11.1 for finding an implicit representation for a parameterized set.

## 10.12 Places of a Curve

This section is concerned with studying the local structure of a plane curve. In particular, we would like to analyze the curve in a neighborhood of a singularity. The analysis will be carried out by expanding the curve locally using power series. Again, if we were to jump right in and begin with all the relevant definitions, readers new to algebraic geometry would probably find everything very abstract and even if they would have no trouble following the technical details it would seem to be a lot of formal mumbo-jumbo. For that reason, we shall start this section with some motivation for what we are going to do. The motivation comes from complex analysis. We will basically take a well-known theory in complex analysis and translate it into an algebraic setting. Therefore, we start with a simple example on the complex analysis side and subsequently show how its analysis has bearing on the analysis of curves in algebraic geometry. This will hopefully clarify some of the issues at stake. We are relying on the fact that every reader has had calculus and probably a little complex analysis, so that the ideas should make a little more sense here and help the reader understand the more algebraic and abstract discussion that follows.

Consider the "function"

$$w = z^{1/2}. \tag{10.65}$$

This w is actually not a function of z because it is not single-valued. Over the **reals** one would normally break it into two functions

$$w_+(z) = +\sqrt{z} \quad \text{and} \quad w_-(z) = -\sqrt{z}, \quad z \in [0,\infty),$$

with their domains restricted to the nonnegative reals. Here √ is the usual real-valued function defined on nonnegative reals that returns the nonnegative square root of its argument. If one allows complex values, a function version of w for negative real z is

$$w_0(z) = +i\sqrt{-z}, \quad z \in (-\infty,0].$$

It is easy to see that the two functions $w_+(z)$ and $w_0(z)$ would then define a nice continuous function defined on all of the reals. Things get more complicated if one considers z as a complex number. In that case, writing z in the polar form

$$z = r\,e^{i\theta}, \quad \text{with} \quad r,\theta \in \mathbf{R} \quad \text{and} \quad 0 \le r,$$

we can express two single-valued functions or "branches" for w in the form

$$w_1(z) = \sqrt{r}e^{i\theta/2}, \tag{10.66a}$$

$$w_2(z) = -\sqrt{r}e^{i\theta/2}. \tag{10.66b}$$

The problem we run into now is one of continuity, or lack of it, because in the complex plane there are many paths from one point to another. For example, one can start at $e^0 = 1$ and then get back to that same point by walking along the points $e^{i\theta}$, $0 \le \theta \le 2\pi$, of the unit circle in a counter clockwise fashion. The function $w_1(z)$ starts out as +1 but approaches −1 at the end. What this basically shows is that it is not possible to define a continuous single-valued square root function on the unit circle.

Returning to the issue of single-valuedness, one could, of course, pick one or the other of the branches and forget about the other one, but this would be an unsatisfactory solution since the multiple-valuedness of the function is an important aspect of it. Therefore, with the domain of functions restricted to the complex plane, one is stuck with a concept of multiple-valued function. On the other hand, we shall see that if one allows enlarging the domain of a function to a "Riemann surface," then one can restore the single-valuedness of functions. Historically, searching for a continuous way to pass from one value of a function at a point to another first lead to the concept of analytic element and analytic continuation, something introduced by K. Weierstrass.

Let us call a pair (f,c) an *analytic element with center c* if f is an analytic function in a neighborhood of c. We know f has a power series expansion

$$f(z) = a_0 + a_1(z-c) + a_2(z-c)^2 + \ldots \tag{10.67}$$

with a certain positive radius of convergence r called the *radius* of the analytic element (f,c). The open disk of radius r about c will be called the *disk* of (f,c). Note that the function f may be analytic over a larger region than its disk but the formula (10.67) is valid only over that disk. How to extend the definition of f? Choose a point d in the disk, $|d - c| < r$. The series in (10.67) can be rearranged into a new power series

$$f_2(z) = b_0 + b_1(z-d) + b_2(z-d)^2 + \dots \tag{10.68}$$

about d. The radius of convergence of the series in (10.68) is at least $r - |c - d|$, but it may be larger. If it is larger then since f and $f_2$ agree on an open set, they define an analytic function with domain larger than that of f. In any case, the analytic element $(f_2,d)$ is called a *direct analytic continuation* of the element (f,c). A point either in or on the boundary of the disk for (f,c) which lies in the interior of the disk of a direct analytic continuation of (f,c) is called a *point of continuability* of (f,c); otherwise, it is called a *point of noncontinuability* or a *singular point* of (f,c). For example, one can show that 0 is a singular point for the analytic elements $(w_i(z),1)$, $i = 1, 2$, where $w_i(z)$ are the square root functions in (10.66).

If we repeat the process of continuation starting with $f_2$, we shall get a sequence of analytic functions $f, f_2, f_3, \dots$. See Figure 10.13. This general process of extending functions by a sequence of direct analytic continuations is called *analytic continuation*. One can continue functions uniquely along curves, but if two curves end up at the same point, the two functions we get at the end of the curves by this process may not be the same. Weierstrass defined a *global analytic function* to be the totality of analytic elements that can be obtained by analytic continuation from a given one. (Actually, Weierstrass did not use the adjective "global". He also dealt with holomorphic, that is, differentiable, functions. Although the terms "holomorphic" and "analytic" are often used interchangeably, "analytic" has a more general connotation and can allow certain singularities.) One can prove that if two global analytic functions have an analytic element in common, then they are identical.

Given two analytic elements $(f_1,c_1)$ and $(f_2,c_2)$ belonging to the same global analytic function, one says that they determine the same *branch* at a point c that belongs to the intersection of their disks if the functions $f_1$ and $f_2$ are identical in a neighborhood of c. Determining the same branch at a point c is an equivalence relation on the set of analytic elements that contain c in their disk. An equivalence class is called an *analytic branch at c*. The equivalence classes are in one-to-one correspondence with power series in $(z - c)$, which have a positive radius of convergence. For example, the expansion of $w_1(z)$ about the point 1 is

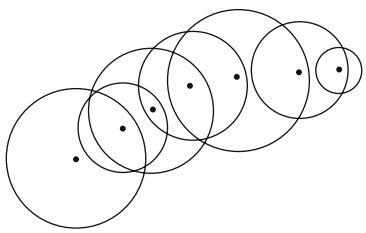


**Figure 10.13.** Analytic continuation.

$$w_1(z) = 1 + \frac{1}{2}(z-1) - \frac{1\cdot 1}{2\cdot 4}(z-1)^2 + \frac{1\cdot 1\cdot 3}{2\cdot 4\cdot 6}(z-1)^3 - \ldots$$

and since $w_2(z) = -w_1(z)$, it has a different series expansion.

Given a global analytic function $F = \{(f,c)\}$, associate to each equivalence class of $(f,c)$ the well-defined value $f(c)$. This defines a well-defined single-valued function on the *Riemann surface*, which is the set of analytic branches of F. The Riemann surface has a natural topology associated to it and is in general a legitimate surface. One can visualize the Riemann surface in the case of the square root function in (10.65) as follows (see Figure 10.14): We superimpose two copies, called sheets, of the complex plane on top of each other and cut each along the positive x-axis starting at the origin. Think of the edge of the cut along the upper half plane side as the upper cut and the other edge as the lower cut on each branch. Then starting at point **A** on the lower sheet (the branch defined by $w_1(z)$) we continue around to the points **B**, **C**, **D**, **E**, and **F** on the same sheet. Leaving **F** we get to the lower cut of the first sheet. There we jump to the upper cut of the second sheet (defined by the function $w_2(z)$) and continue on to **G** on the second sheet. Then move on to **H**, **I**, and **J** on the second sheet. From **J** we reach the lower cut of the second sheet. At that point we jump back to the upper cut of the first sheet and arrive back at **A**. The positive axis we cut along is called a *branch line* and the origin is called a *branch point*. Actually, we could have cut along any curve that starts at 0 and goes to infinity.

Because we are dealing with multiple-valued functions, rather than expressing them in the functional notation $w = f(z)$, it is more convenient to think of them and their associated Riemann surface as defined implicitly by the equation

$$w - f(z) = 0. \tag{10.69}$$

The Riemann surface is actually a complex manifold, meaning that its coordinate neighborhoods are defined by analytic functions (rather than just differentiable real-valued functions). Basically, this means that every point on the surface has a neighborhood where we can parameterize both variables z and w by a single (complex) parameter t in the form

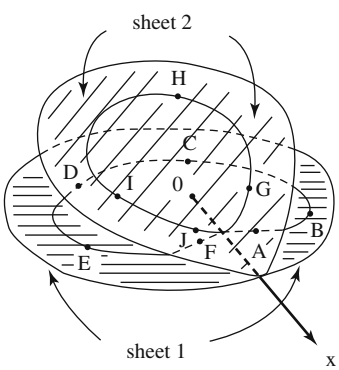



**Figure 10.14.** The Riemann surface for $z^{1/2}$.

$$z = g_1(t), \quad w = g_2(t) = f(g_1(t)), \tag{10.70}$$

where $g_1(t)$ and $g_2(t)$ have power series expansions in t that have a positive radius of convergence. Such a variable t is called a (local) *uniformizing variable* and the process is called *uniformization*. It exists because of the very definition of a Riemann surface in terms of analytic elements and convergent power series. An expansion like in (10.67) translates into a uniformization

$$z = c + t, \quad w = a_0 + a_1 t + a_2 t^2 + \dots,$$

with uniformizing variable t. For example, the square root in (10.65) is replaced by an equation

$$w - z^{1/2} = 0 \tag{10.71}$$

and happens to have an especially simple uniformization

$$w = t, \quad z = t^2. \tag{10.72}$$

The uniformization in (10.72) is actually valid for the whole Riemann surface associated to (10.71) and not just to a neighborhood of 0. That is not always possible. Whether or not a global uniformization exists for a Riemann surface is called the *uniformization problem* in complex analysis. The interested reader can find a nice discussion of this problem and its history in [Abik81].

The pair of functions $(g_1(t), g_2(t))$ in equations (10.70) is just another representation of what we called an analytic element. As we just pointed out, from a topological point of view it simply corresponds to showing that the Riemann surface is a surface. Thinking of things in that way, since there are many ways to coordinatize a neighborhood about a point in a manifold, why not try to find a special coordinate neighborhood that gives us more information about this neighborhood. Define two representations $(g_1(t), g_2(t))$ and $(h_1(t), h_2(t))$ to be *equivalent* if there exists a power series

$$u(t) = c_1 t + c_2 t^2 + \dots, \quad c_1 \neq 0, \tag{10.73}$$

which converges in a neighborhood of 0, and is one-to-one there, so that

$$h_1(t) = g_1(u(t)),$$
$$h_2(t) = g_2(u(t)).$$

One can prove that this is an equivalence relation and that if z has an expansion

$$z = a_0 + a_m t^m + a_{m+1} t^{m+1} + \dots \tag{10.74}$$

with $m > 0$ and $a_m \neq 0$, then z has an equivalent expansion

$$z = a_0 + b_m s^m. \tag{10.75}$$

In fact, if the representation (10.75) was obtained by substituting the series u(t) in (10.73) for t in (10.74), then (up to equivalence) all the other different representation can be obtained by substituting $u(\omega t)$ for t in (10.74), where $\omega$ is an arbitrary mth root of unity. The point $a_0$ is called a point of *order* m. If m = 1, then $a_0$ is called a *regular* point. It follows that we are either at a regular point or our surface looks like

$$w = z^{1/m}.$$

This finishes our quick look at analytic continuation and Riemann surfaces and how they play a role in the analysis of complex functions and their singularities. To keep things as simple as possible we only dealt with holomorphic functions, but in any serious discussion one would also consider meromorphic functions so that quotients of power series would replace the power series in the discussion above. This will show up later on when we use quotient fields rather than rings.

We now move on to an example that brings out the relevance of what was just discussed to algebraic geometry. Consider the affine plane curve **C** in $\mathbf{C}^2$ defined by

$$y^2 = x^2(1 - x). \tag{10.76}$$

**Note:** It turns out that for every irreducible cubic curve with an ordinary double point we can find an affine coordinate system so that in that coordinate system the curve is defined by the equation (10.76) (see [BriK81]).

The curve **C** has a singularity at the origin, namely, a double point. See Figure 10.15. We can parameterize the curve by computing the intersection of the lines y = xt with the curve. This will give the parameterization

$$c(t) = (x(t), y(t)),$$

where

$$x(t) = 1 - t^2$$
$$y(t) = t - t^3.$$

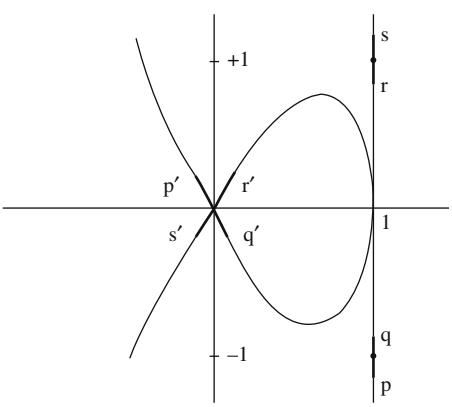



**Figure 10.15.** Parameterizing $y^2 = x^2(1 - x)$.

We can think of this map as sending the complex line x = 1 onto the complex curve with t = +1, −1 mapping onto the origin. The map is onto and one-to-one except for these two points. The line x = 1 is obviously a curve with no singularities. Later on we shall look at this as an example of how singularities are "resolved." Although c is not globally one-to-one, it is locally so. Small enough neighborhoods of +1 get mapped in a one-to-one fashion onto a set in **C**, similarly for a small neighborhood of −1. See Figure 10.15 again, where the points **p**, **q**, **r**, and **s** get mapped onto **p′**, **q′**, **r′**, and **s′**, respectively. One can think of this as having broken the curve **C** into more primitive parts, in analogy with how we factored polynomials into their irreducible factors and expressed varieties as unions of irreducible varieties. In our case the polynomial

$$f(X,Y) = Y^2 - X^2(1-X)$$

in the ring **C**[X,Y] that defines **C** is irreducible. But then again, we are now dealing with a local decomposition of the curve, not a global one as in the case of factoring polynomials. Is there another ring in which f factors? Factoring into polynomials did not work. Factoring into rational functions would also not work. However, suppose we consider holomorphic functions in two variables in a neighborhood **U** of the origin. In this case we **can** factor f as $f = f_1 f_2$, where

$$f_1(x,y) = y + x\sqrt{1-x},$$
$$f_2(x,y) = y - x\sqrt{1-x},$$

and we have used one of the branches of the square root function. To avoid the problem with the neighborhood **U** not being well defined, one passes to equivalence classes of holomorphic functions. In other words, we have a factorization in the ring of holomorphic branches at the origin, where "branch" is being used in the sense described above. We have found a ring of the type we were looking for. Singularities of curves can be detected by checking the "irreducibility" of the curve in this ring.

We are done with the introductory remarks for this section. See [BriK81] for a much more complete discussion of how complex analysis comes into the picture for algebraic geometry. Being part of analysis, the methods and approaches described above make heavy use of convergence and differentiability, concepts intimately connected with the topology of the complex plane. Topology also plays a role in algebraic geometry, but algebraic geometers prefer not to have to worry about convergence of series and approach the same questions in a purely algebraic way. They are guided by what is known from analysis, but replace convergent series with formal power series where convergence is not an issue. Hopefully, this abstract algebraic approach, which we shall now describe, will make more sense with our background discussion.

We could start with the most concrete case of an affine plane curve defined by an equation

$$f(X,Y) = 0.$$

The plan would then be to express X and Y as formal power series in a variable t. This is plausible because an implicit function type theorem basically states that this is pos-

sible in the world of continuous functions and convergent series, but we should not expect this representation to be unique. One reason for the nonuniqueness can be explained by the that fact that the curve is derived from a projective curve that can be coordinatized in many ways. For this and other reasons it is better to start with a projective curve from the beginning. Our discussion will follow the presentation of places of curves as given in [Walk50].

Let **C** be a plane curve in $\mathbf{P}^2(k)$. As usual, because it is convenient to study the curve in the context of a coordinate system, we pick one, but one of our tasks as we go along is to make sure that everything we do is independent of such a choice. Assume that the curve **C** is defined by an equation

$$F(X, Y, Z) = 0,$$

where F(X,Y,Z) a homogeneous polynomial in our coordinate system.

**Definition.** A (*projective*) *parameterization* of **C** (with respect to the given coordinate system) is a triple (X(t),Y(t),Z(t)) of rational functions X(t), Y(t), Z(t) $\in$ k((t)) satisfying

(1) F (X(t),Y(t),Z(t)) = 0.
(2) For no h(t) $\in$ k((t)) do all of the products h(t)X(t), h(t)Y(t), or h(t)Z(t) belong to k.

From a technical point of view, to make sense of some statements we should think of our curve as a curve in $\mathbf{P}^2(k((t)))$, the projective plane over the field k((t)). For example, condition (1) only really makes sense in that context, namely, we think of F as a polynomial over k((t)). We shall not emphasize this point to keep the discussion simple. All the important concepts will live in $\mathbf{P}^2(k)$. Also, condition (2) can be rephrased as saying that [X(t),Y(t),Z(t)] is a point of $\mathbf{P}^2(k((t)))$ that does not lie in $\mathbf{P}^2(k)$. It ensures that a parameterization deals with nonconstant, nontrivial rational functions. We could pick representatives X(t), Y(t), and Z(t) that are formal power series, but this will not always be possible when we switch to affine coordinates. The field k((t)) is the algebraic analog of the meromorphic functions in analysis. What we call a parameterization here is sometimes called a *branch representation* (see [Seid68]).

Given how coordinate transformations are defined, we leave it to the reader to check that a parameterization defined in one coordinate system will be a parameterization (satisfying (1) and (2)) when transformed to another coordinate system (Exercise 10.12.1). This means that a parameterization defined in one coordinate system defines a well-defined parameterization in any other coordinate system. We shall allow ourselves to talk about "the parameterization (X(t),Y(t),Z(t))" even though strictly speaking we shall mean [X(t),Y(t),Z(t)]. Furthermore, since it is the equivalence class [X(t),Y(t),Z(t)] that is important, we shall feel free to switch to whatever representative (X(t),Y(t),Z(t)) is convenient.

Let (X(t),Y(t),Z(t)) be a parameterization of **C** and let

$$m = -\min(\mathrm{ord}(X(t)), \mathrm{ord}(Y(t)), \mathrm{ord}(Z(t))).$$

By multiplying each coordinate by $t^m$ we may assume that our parameterization has the property that each coordinate is a formal power series

$$X(t) = a_0 + a_1 t + a_2 t^2 + \dots,$$
$$Y(t) = b_0 + b_1 t + b_2 t^2 + \dots,$$
$$Z(t) = c_0 + c_1 t + c_2 t^2 + \dots,$$

and at least one of ord(X(t)), ord(Y(t)), or ord(Z(t)) is zero. It follows that

$$c = (X(0), Y(0), Z(0)) = (a_0, b_0, c_0) \neq (0,0,0).$$

**Definition.** The point [c] in $\mathbf{P}^2(k)$ is called the *center* of the parameterization.

**10.12.1. Proposition.** The center of a parameterization for a plane curve is independent of the coordinate system and lies on the curve.

**Proof.** The proof of the first part is straightforward and left as an exercise. One must show that related parameterizations define the same center. Although the proof of the second part is not hard, the fact may seem more obvious than it actually is. The problem is that we are dealing with infinite power series and evaluation of such has to be handled carefully. Specifically, what we know is that

$$F(X(t), Y(t), Z(t)) = 0, \tag{10.77}$$

but all we know immediately is that F(X(t),Y(t),Z(t)) is a power series G(t). There is no à priori relationship between G(0) and F(c). One way to prove that [c] lies on the curve is to develop a theory of congruence first (something that is needed in Theorem B.7.5 to prove that substitution into power series is alright). Equation (10.77) then implies

$$F(X(t), Y(t), Z(t)) \equiv 0 \pmod{t}.$$

This equation and the fact that

$$X(t) \equiv X(0) \pmod{t},$$
$$Y(t) \equiv Y(0) \pmod{t},$$
$$Z(t) \equiv Z(0) \pmod{t},$$

imply that F(c) = 0 (mod t), from which the result follows.

Next, given an arbitrary parameterization (X(t),Y(t),Z(t)) of $\mathbf{C}$, let $h(t) \in k[[t]]$, where $h \neq 0$ and $\mathrm{ord}(h) > 0$. Let

$$X_h(t) = X(h(t)), \quad Y_h(t) = Y(h(t)), \quad \text{and} \quad Z_h(t) = Z(h(t)).$$

It is easy to show that

**10.12.2. Proposition.** $(X_h(t), Y_h(t), Z_h(t))$ is a parameterization with the same center as that of (X(t),Y(t),Z(t)).

**Proof.** Exercise.

**Definition.** If ord(h) = 1, then we say that the parameterizations $(X_h(t),Y_h(t),Z_h(t))$ and $(X(t),Y(t),Z(t))$ are *equivalent*.

**10.12.3. Proposition.** Being equivalent is an equivalence relation on the set of parameterizations for a plane curve. Equivalent parameterizations have the same center.

**Proof.** Exercise.

Another way of looking at the relation "equivalent" is that it corresponds to an automorphism of k[[t]] over k.

**Definition.** A parameterization of the form $(X(t),Y(t),Z(t))$, with $X(t), Y(t), Z(t) \in k[[t^m]]$ and $m > 1$, is called *reducible*; otherwise it is called *irreducible*.

We shall be interested in irreducible parameterization because reducible ones can be simplified using the substitution $s = t^m$.

Up to now we have used homogeneous coordinates. Now let us translate everything to affine coordinates. Assume that $Z(t) \neq 0$ in a parameterization $(X(t),Y(t),Z(t))$. Let

$$f(X,Y) = F(X,Y,1)$$

be the affine equation of the curve **C**. We can think of

$$\overline{X}(t) = \frac{X(t)}{Z(t)}$$
$$\overline{Y}(t) = \frac{Y(t)}{Z(t)}$$

as an affine parameterization. It satisfies

$$f(\overline{X}(t),\overline{Y}(t)) = 0. \tag{10.78}$$

Conversely, given any $\overline{X}(t)$ and $\overline{Y}(t)$ satisfying equation (10.78), we get a projective parameterization $(\overline{X}(t),\overline{Y}(t),1)$.

Note that in the interesting special case where ord(Z(t)) = 0 it follows from Theorem B.11.13 that 1/Z(t) belongs to k[[t]] so that both $\overline{X}(t)$ and $\overline{Y}(t)$ also belong to k[[t]]. Since we can always assume that one of ord(X(t)), ord(Y(t)), or ord(Z(t)) is zero, there is always one coordinate system with respect to which the affine curve is parameterized by formal power series.

**Definition.** An (*affine*) *parameterization* of **C** (with respect to a given coordinate system) is a pair $(X(t),Y(t))$ of rational functions $X(t), Y(t) \in k((t))$ satisfying

(1) $f(X(t),Y(t)) = 0$. (Here, f is thought of as a polynomial over the field $k((t))$).
(2) Not both $X(t)$ and $Y(t)$ belong to k.

The notions of equivalent and reducible or irreducible parameterizations carry over to the affine ones in the obvious way.

**10.12.4. Theorem.** Given any parameterization we can always find a coordinate system so that in that coordinate system the related affine parameterization is equivalent to one of the form

$$X(t) = t^m$$
$$Y(t) = a_1 t^{m_1} + a_2 t^{m_2} + \dots,$$

where $a_1 \neq 0$, $0 < m$, and $0 < m_1 < m_2 < \dots$

**Proof.** See [Walk50].

The next theorem gives us a criterion for when a parameterization is reducible.

**10.12.5. Theorem.** An affine parameterization of the form shown in Theorem 10.12.4 is reducible if and only if the integers $m_1, m_2, \dots$ have a common factor larger than 1.

**Proof.** See [Walk50].

**Definition.** A *place* of a plane curve is an equivalence class of irreducible parameterizations of the curve with respect to being equivalent. The point on the curve determined by the center of any one the representatives of a place is called the *center* of the place.

A place is the algebraic version of what is called a branch of a function in complex analysis and for that reason is sometimes also called a *branch*. Proposition 10.12.3 shows that the center of a place is well defined. One can talk about a place of a projective curve or any of its associated affine curves. One always talks about them in the context of a particular representative for its equivalence class.

Proposition 10.12.1 proved that the center of a place is a point on the curve. At this point, however, we do not know if any places or parameterizations even exist. To put it another way, can curves be represented locally by power series? Therefore, the next theorem is fundamental to the subject.

**10.12.6. Theorem.** Every point of a plane curve in $\mathbf{C}^2$ is the center of at least one but no more than a finite number of places.

Theorem 10.12.6 can be proved in different ways. We follow [Seid68] because this approach will introduce certain local quadratic transformations which are useful for computations. We start with a sequence of lemmas. Also, it will be convenient to work with an affine representation of a curve in the rest of this section.

**10.12.7. Lemma.** If the origin is a simple point of a plane curve $f(X,Y) = 0$ and if the tangent is not vertical there, then there exists a unique power series

$$h(X) = c_1X + c_2X^2 + \ldots$$

so that

$$f(X,h(X)) = 0.$$

**Proof.** To prove the existence of $h(X)$ we shall inductively construct polynomials

$$h_i(X) = c_1X + \ldots + c_iX^i,$$

so that $\operatorname{ord}(f(X,h_i(X))) > i$. Now, since the origin lies on the curve and there is no vertical tangent there, we may assume that

$$f(X,Y) = aX + Y + (\text{terms of order} > 1).$$

for some constant a. Let $c_1 = -a$ and $h_1(X) = -aX$. Assume that $h_{i-1}$ has been defined for $i > 1$. Using the Taylor series expansion for f about the "point" $(X,c_1X + \ldots + c_{i-1}X^{i-1})$ (see Equation (10.39)) and the fact that

$$\frac{\partial f}{\partial Y} = 1 + g(X,Y), \quad \text{where ord } g > 0,$$

we see that

$$\begin{aligned} f\left(X,c_1X + \ldots + c_{i-1}X^{i-1} + c_iX^i\right) &= f\left(X,c_1X + \ldots + c_{i-1}X^{i-1}\right) \\ &\quad + \left(1 + g\left(X,c_1X + \ldots + c_{i-1}X^{i-1}\right)\right)c_iX^i \\ &\quad + (\text{terms of degree} > i) \\ &= \left(dX^i + (\text{terms of degree} > i)\right) + c_iX^i \\ &\quad + (\text{terms of degree } > i), \end{aligned}$$

for some d. Therefore, let $c_i = -d$. This sequence of $c_i$ clearly guarantees that

$$f\left(X,c_1X + c_2X^2 + \ldots\right) = f\left(X,c_1X + \ldots + c_{i-1}X^{i-1}\right) + X^i(\ldots),$$

which implies that

$$\operatorname{ord} f\left(X,c_1X + c_2X^2 + \ldots\right) \geq i, \quad \text{for all } i,$$

and so $f(X,c_1X + c_2X^2 + \ldots)$ must be identically zero. This proves the existence of the desired power series. To prove uniqueness one assumes that there are two power series satisfying the hypothesis of the lemma and then inductively shows that their ith coefficients must be equal. See [Seid68].

**10.12.8. Theorem.** Every simple point of a plane curve in $\mathbf{C}^2$ is the center of a unique place.

**Proof.** By translating and rotating the curve if necessary, we may assume that the simple point is the origin and that it has a nonvertical tangent. Let h(X) be the power series guaranteed by Lemma 10.12.7 and set

$$u_0(t) = t \quad \text{and} \quad v_0(t) = c_1 t + c_2 t^2 + \ldots$$

using the notation of that Lemma. It follows that $(u_0(t),v_0(t))$ determines a place of the curve with center the origin. On the other hand, let (u(t),v(t)) be an arbitrary irreducible parameterization of the curve with center the origin. By Theorem 10.12.4 we may assume that ord(u), ord(v) > 0. Now

$$f(X, c_1X + c_2X^2 + \ldots) = 0$$

implies that

$$f(X,Y) = (Y - c_1X - c_2X^2 - \ldots)g(X,Y)$$

for some polynomial g(X,Y) in k[[X]][Y]. The polynomial g(X,Y) has a nonzero constant term since f has linear terms. Therefore,

$$f(u,v) = (v - c_1u - c_2u^2 - \ldots)g(u,v) = 0$$

in k[[t]]. Because g(X,Y) has a nonzero constant term it is a unit in k[[t]]. Therefore we must have

$$v = c_1u + c_2u^2 + \ldots,$$

which implies that ord v = 1, since the parameterization is irreducible. It follows that (u(t),v(t)) and $(u_0(t),v_0(t))$ are equivalent parameterizations and hence determine the same place. This proves the theorem.

Next, we handle singular points **p** of a plane curve. The idea will be to transform the curve to a new curve so that if **q** is a point on the new curve corresponding to the point **p** on the original curve, then **q** is no longer singular. By Theorem 10.12.8 the new curve will have a place with center **q** and this place can be transformed back to a place with center **p** on the original curve. As an example of the kind of transformation we have in mind, consider

$$U = X, \quad V = \frac{Y}{X}. \tag{10.79}$$

**Definition.** The transformation defined by equations (10.79) is called a *(local) quadratic transformation* with *center* (0,0).

First of all, here are some simple observations about the map in (10.79). As a mapping from the plane to the plane, it is of course not defined for points on the y-axis; however, consider a line Y = mX. The transformation (10.79) will send the points of that line other than **0** to the line V = m. See Figure 10.16. What this means is that if the origin is a singular point for a plane curve **C** and if the tangent lines to the curve at the origin have slopes $m_1, m_2, \ldots, m_r$, then $(0,m_i)$ will be the points on the transformed curve **C′** corresponding to the origin on **C** and **C′** will have tangent lines V =

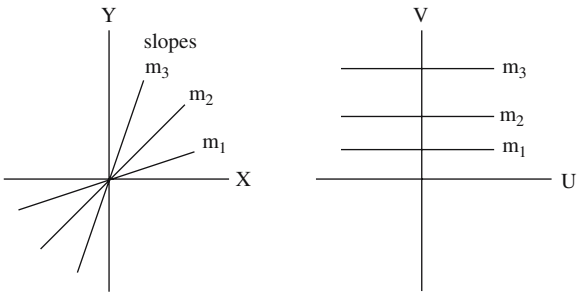



**Figure 10.16.** An example of the transformation defined by (10.79).

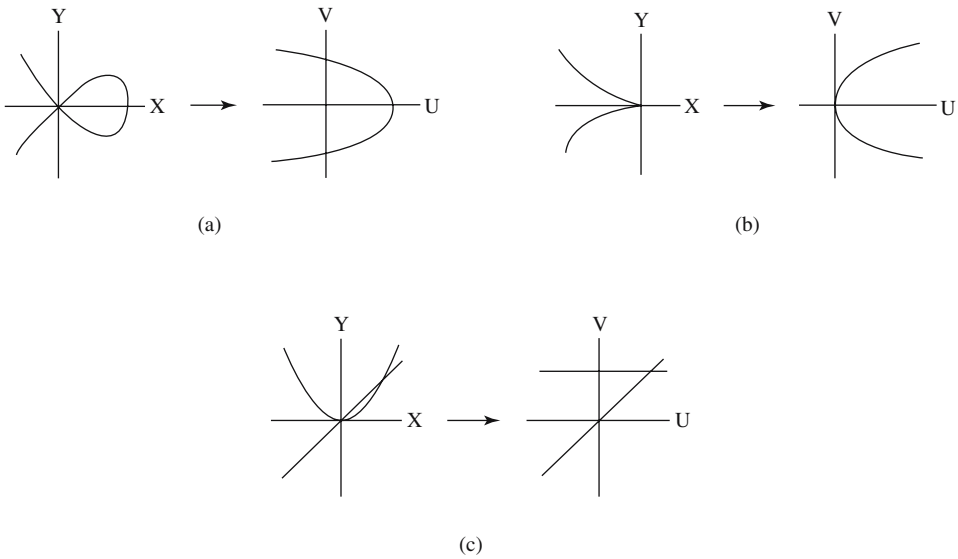



**Figure 10.17.** Quadratic Transformations of Curves.

$m_i$ at those points. If the $m_i$ are distinct, then in effect we have separated the singularity into distinct points that are now simple points (at least if the original tangent lines had multiplicity 1). This process is referred to as *blowing up* or *resolving* the singularity.

**10.12.9. Example.** $X^3 - X^2 + Y^2 = 0$

**Result.** Equations (10.79) transform this into $U - 1 + V^2 = 0$. See Figure 10.17(a).

**10.12.10. Example.** $X^3 - Y^2 = 0$

**Result.** Equations (10.79) transform this into $U - V^2 = 0$. See Figure 10.17(b).

**10.12.11. Example.** $(Y - X^2)(Y - 3X) = 0$

**Result.** Equations (10.79) transform this into $(V - U)(V - 3) = 0$. See Figure 10.17(c).

Let **C** be an affine plane curve with equation

$$f(X, Y) = 0$$

and let $(X(t), Y(t))$ be a parameterization of **C** with center **0**.

**Definition.** The parameterization $(X(t), Y(t)/X(t))$ will be called the *transformed parameterization* of **C** with respect to the quadratic transformation (10.79).

Write

$$f(X, Y) = f_r(X, Y) + f_{r+1}(X, Y) + \dots,$$

where the $f_i(X,Y)$ are homogeneous polynomials of degree i and $f_r(X,Y) \neq 0$. Then

$$\begin{aligned} f(X, XY) &= f_r(X, XY) + f_{r+1}(X, XY) + \dots \\ &= X^r(f_r(1, Y) + Xf_{r+1}(1, Y) + \dots), \end{aligned}$$

which shows that $f(X,XY)$ is divisible by $X^r$ and no higher power of X. Let $\mathbf{C}_v$ be the affine curve defined by

$$f_v(X, Y) = \frac{f(X, XY)}{X^r}.$$

**Definition.** The plane curve $\mathbf{C}_v$ is called the *transformed curve* with respect to the quadratic transformation (10.79).

**10.12.12. Lemma.** If **C** is irreducible and not the line $X = 0$, then $\mathbf{C}_v$ is irreducible.

**Proof.** We need to show that $f_v(X,Y)$ is irreducible.

**Claim.** $f_v(X,Y)$ is not divisible by a nonconstant polynomial $g(X)$.

If $f_v(X,Y) = g(X)h(X,Y)$, then replacing Y by Y/X and multiplying through by $X^r$ would mean that $f(X,Y) = X^r g(X)h(X,Y/X)$ and the irreducibility of f would imply that $g(X) = cX^d$ for some constant c. But $f_v(X,Y)$ is not divisible by X and so $d = 0$ and the claim is proved.

Now assume that $f_v(X,Y)$ is a product of two polynomials. By the claim, these polynomials would contain terms with Y to a positive power. Again replacing Y by Y/X and multiplying through by an appropriate power of X would show that $X^d f(X,Y)$ is a product of two polynomials, which is impossible. The lemma is proved.

**10.12.13. Lemma.** If **C** is a plane curve that does not have a vertical tangent at **0**, then there is a one-to-one correspondence between the parameterizations of **C** centered at **0** and the parameterizations of $\mathbf{C}_v$ centered on the finite y-axis.

**Proof.** See [Seid68]. The transformation (10.79) sets up the desired correspondence between parameterizations.

The next lemma is another special case of Theorem 10.12.6.

**10.12.14. Lemma.** An ordinary r-fold point **p** of **C** has exactly r parameterizations centered at **p** that are linear and that have the same tangents as those of **C**.

**Proof.** We may assume that the r-fold point, $r > 0$, is the origin and, since our base field is the complex numbers,

$$f(X,Y) = \prod_{i=1}^{r}(Y - m_i X) + \dots .$$

The transformed curve $\mathbf{C}_v$ will be defined by

$$f_v(X,Y) = \prod_{i=1}^{r}(Y - m_i) + X(\dots).$$

It follows that $X = 0$ intersects $\mathbf{C}_v$ precisely in the points $(0,m_i)$. Furthermore, the multiplicity of the intersection points is 1 because all the $m_i$ are distinct. This means that the points $(0,m_i)$ are simple points for $\mathbf{C}_v$ and that $\mathbf{C}_v$ has r parameterizations centered on $X = 0$. We conclude that **C** has r branches centered at the origin. From Lemma 10.12.7 and Theorem 10.12.8, the parameterization of $\mathbf{C}_v$ at $(0,m_i)$ is given by

$$\begin{aligned} X &= t \\ Y &= m_i t + \dots \end{aligned}$$

and its tangent line is $Y = m_i X$. The lemma is proved.

**10.12.15. Lemma.** Any irreducible curve can be transformed into a curve with only ordinary singularities with a finite sequence of quadratic transformations.

**Proof.** See [Walk50].

After these preliminaries, we are ready to prove Theorem 10.12.6.

**Sketch of proof for Theorem 10.12.6.** Lemma 10.12.7 and Lemma 10.12.14 already proved two special cases. Let **p** be an arbitrary r-fold point of **C**. We may assume that **C** is irreducible, $\mathbf{p} = \mathbf{0}$, and $X = 0$ is not a tangent of **C** at **0**. Since our base field is the complex numbers,

$$f(X,Y) = \prod_{i=1}^{r}(Y - m_i X) + \dots ,$$

where the $m_i$ are not necessarily distinct. It follows that

$$f_v(X,Y) = \prod_{i=1}^{r}(Y - m_i) + X(\ldots),$$

which shows that the sum of the intersection multiplicities of the line X = 0 with the curve $\mathbf{C}_v$ is r.

Now if two of the $m_i$ are distinct, then, since the intersection multiplicities at those points would be each less that r, we could prove this case of Theorem 10.12.6 using induction on r. But what if all the $m_i$ are equal? We would have

$$f_v(X,Y) = (Y - m_1)^r + X(\ldots).$$

We can now translate $\mathbf{C}_v$ so as to move the point $(0,m_1)$ to the origin. The new curve does not have a vertical tangent at the origin, so that we can apply a quadratic transformation to it. Unfortunately, the new curve could also have an r-fold singularity with multiplicity r. The question arises, if we keep repeating this process of translating to the origin and applying a quadratic transformation, will we continue to get r-fold singularities at the origin with multiplicity r? The answer is no. See [Seid68] for the details. Therefore, after a finite number of steps our procedure will lead us to a situation where we have a simple point. Induction works.

Places provide an alternate way to prove various theorems of algebraic geometry such as Bèzout's theorem. See [Walk50]. Some definitions can also be phrased in terms of places. For example,

**10.12.16. Theorem.** A point of a plane curve is nonsingular if it is the center of just one linear place.

**Proof.** See [Walk50]. The only if part of this theorem is Theorem 10.12.8.

The definition of a place given in this section was chosen because it is less abstract than other definitions. However, it is worth being aware of another common definition that is used in "valuation theory." Let $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in k^n$. Then $\mathbf{c}$ defines a homomorphism

$$\begin{aligned} v:\ & k[X_1, X_2, \ldots, X_n] \to k \\ & f(X_1, X_2, \ldots, X_n) \to f(\mathbf{c}) = f(c_1, c_2, \ldots, c_n). \end{aligned}$$

The element $f(\mathbf{c})$ is called a *value* of f. Now $k[X_1, X_2, \ldots, X_n]$ is a subring of the quotient field $K = k(X_1, X_2, \ldots, X_n)$ and v extends to a homomorphism on K except at those elements whose denominators vanish on $\mathbf{c}$. This motivates the following:

Let K be a field. A subring R of K is called a *valuation ring* if for all $x \in K$, either x or 1/x belongs to R. A *place* for K is a nonzero homomorphism p from a valuation ring $K_p$ in K to a field F, so that if $x \in K$ and $x \notin K_p$, then $p(1/x) = 0$.

The new definition of a place leads to a development of results in algebraic geometry that parallels what we can do with our definition. For more on this approach see Volume II of [ZarS60]. Valuation rings are also discussed in [Kend77].

## 10.13 Rational and Birational Maps

The material in this section is necessary background material for what we have to say about when an implicitly defined plane curve can be parameterized in the next section. We start with an example.

**10.13.1. Example.** Consider the affine conic **C** defined by

$$f(X,Y) = X^2 - XY + Y^2 - 3X = 0. \tag{10.80}$$

See Figure 10.18. The point $\mathbf{p}_0 = (x_0,y_0) = (4,2)$ lies on **C**. The line **L** through $\mathbf{p}_0$ with slope t has equation

$$Y = 2 + t(X - 4). \tag{10.81}$$

To find the intersections of **L** with **C**, we simply need to substitute the right hand side of equation (10.81) into (10.80) and solve for X. We already have one intersection of **L** with **C** and it is easy to check that the second intersection $(x_1,y_1)$ is given by

$$x_1 = \frac{1 - 4t + 4t^2}{1 - t + t^2},$$
$$y_1 = \frac{2 - 5t + 2t^2}{1 - t + t^2}.$$

The approach used in Example 10.13.1 works to parameterize any nondegenerate affine conic curve **C** defined by

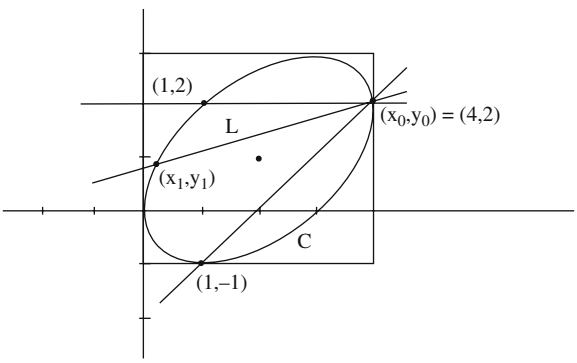



**Figure 10.18.** The conic defined by equation (10.80).

$$f(X,Y) = 0.$$

Assume that $\mathbf{p}_0 = (x_0,y_0)$ lies on $\mathbf{C}$. The line $\mathbf{L}$ through $\mathbf{p}_0$ with slope t has equation

$$Y = y_0 + t(X - x_0). \tag{10.82}$$

We need to solve

$$f(X, y_0 + t(X - x_0)) = 0 \tag{10.83}$$

for X. Rather than using the standard quadratic formula which would seem to involve square roots, note that we already have one solution $x_0$. It follows that the second root $x_1$ satisfies

$$x_1 + x_0 = -B,$$

where B is the coefficient of X in equation (10.83) **after** one has divided the equation by the coefficient of the $X^2$ term. Therefore, the second intersection $(x_1,y_1)$ of $\mathbf{L}$ with $\mathbf{C}$ is determined by rational functions in t since one can use equation (10.82) to solve for $y_1$.

Because conics can be described by rational functions, they are called "rational" curves. Before we introduce the terminology required to study rational functions, let us look at the simpler case of polynomial functions.

**Definition.** Let $\mathbf{V} \subseteq k^n$ and $\mathbf{W} \subseteq k^m$. A function

$$u: \mathbf{V} \to \mathbf{W}$$

is called a *polynomial function from* $\boldsymbol{V}$ *to* $\boldsymbol{W}$ if there exist polynomials $p_1, p_2, \ldots, p_m \in k[X_1,X_2, \ldots ,X_n]$ such that

$$u(\mathbf{a}) = (p_1(\mathbf{a}), p_2(\mathbf{a}), \ldots, p_m(\mathbf{a})), \quad \text{for all } \mathbf{a} \in k^n.$$

The polynomials $p_i$ are called *representatives* for the function u. If m = 1 and $\mathbf{W} = k$, then u will be called simply a *polynomial* or *regular function on* $\boldsymbol{V}$. The set of polynomial functions

$$u: \mathbf{V} \to k$$

on $\mathbf{V}$ will be denoted by $k[\mathbf{V}]$.

The representatives $p_i$ of a polynomial function are typically not unique. For example, if $\mathbf{V}$ is a hypersurface V(f), then $p_i$ and $p_i + f$ define the same function on $\mathbf{V}$ because f vanishes on $\mathbf{V}$. Note that pointwise addition and multiplication makes $k[\mathbf{V}]$ into a ring.

**10.13.2. Theorem.** Let $\mathbf{V} \subseteq k^n$.

(1) $k[\mathbf{V}]$ is isomorphic to $\dfrac{k[X_1,X_2,\ldots,X_n]}{I(\mathbf{V})}$.

(2) $\mathbf{V}$ is irreducible if and only if $k[\mathbf{V}]$ is an integral domain.
(3) Assume that k is algebraically closed. If $\mathbf{V}$ is a hypersurface $V(f)$, then the ring $k[\mathbf{V}]$ is isomorphic to the ring $\dfrac{k[X_1, X_2, \ldots, X_n]}{\sqrt{<f>}}$. If f is irreducible, then $k[\mathbf{V}]$ is an integral domain.

**Proof.** Consider the map

$$\varphi: k[X_1, X_2, \ldots, X_n] \to k[\mathbf{V}],$$

which sends a polynomial to the function it induces. It is an easy exercise to show that $\varphi$ is a ring homomorphism with kernel $I(\mathbf{V})$, which proves (1). The proof of (2) is Exercise 10.13.2 (see [Shaf94]). The first part of (3) follows from the fact that $I(\mathbf{V}) = \sqrt{<f>}$ and the second part follows from (2).

**Definition.** If $\mathbf{V}$ is an affine variety, then $k[\mathbf{V}]$ is called the *coordinate ring* or *ring of polynomial functions* of $\mathbf{V}$.

Because of Theorem 10.13.2(1) we shall feel free to identify $k[\mathbf{V}]$ with

$$\frac{k[X_1, X_2, \ldots, X_n]}{I(\mathbf{V})}.$$

In fact, we will think of it as a vector space over k by identifying the constant functions with k.

To every affine variety $\mathbf{V}$ we have now associated a ring $k[\mathbf{V}]$, the coordinate ring. How does this ring behave with respect to maps?

**Definition.** Let $u: \mathbf{V} \to \mathbf{W}$ be a polynomial function between affine varieties $\mathbf{V}$ and $\mathbf{W}$. Define

$$u^*: k[\mathbf{W}] \to k[\mathbf{V}]$$

by

$$u^*(f) = f \circ u.$$

The map $u^*$ is sometimes called the *pullback map* defined by u.

**10.13.3. Theorem.** $u^*$ is a well-defined ring homomorphism that is the identity on the constant functions. Conversely, if $h: k[\mathbf{W}] \to k[\mathbf{V}]$ is any ring homomorphism that is the identity on the constant functions, then there exists a unique polynomial map $u: \mathbf{V} \to \mathbf{W}$, such that $h = u^*$.

**Proof.** The first part of the theorem is easy. It is the proof of the converse part that is interesting. Assume that $\mathbf{V} \subseteq k^n$ and $\mathbf{W} \subseteq k^m$. The coordinates of $k^m$ induce m coordinate functions

$$y_i : \mathbf{W} \to k.$$

It follows that $h(y_i)$ will be represented by some polynomial $a_i(X_1, X_2, \ldots, X_n) \in k[X_1, X_2, \ldots, X_n]$. One needs to show that the map

$$u : \mathbf{V} \to \mathbf{W}$$

defined by

$$u(x_1, x_2, \ldots, x_n) = (a_1(x_1, x_2, \ldots, x_n), a_2(x_1, x_2, \ldots, x_n), \ldots, a_m(x_1, x_2, \ldots, x_n))$$

is the unique map that does the job. For more details see [CoLO97].

A fact that will be useful later is

**10.13.4. Proposition.** Let $f : \mathbf{V} \to \mathbf{W}$ be a polynomial function between affine varieties. Then $f(\mathbf{V})$ is dense in $\mathbf{W}$ with respect to the Zariski topology if and only if $f^* : k[\mathbf{W}] \to k[\mathbf{V}]$ is one-to-one.

**Proof.** See [CoLO97]. The "only if" part is easy.

**Definition.** Let $\mathbf{V} \subseteq k^n$ and $\mathbf{W} \subseteq k^m$ be affine varieties. A polynomial function

$$u : \mathbf{V} \to \mathbf{W}$$

is called an *isomorphism* if u has an inverse that is also a polynomial function. In that case the two varieties are said to be *isomorphic*.

**10.13.5. Theorem.** Two affine varieties $\mathbf{V} \subseteq k^n$ and $\mathbf{W} \subseteq k^m$ are isomorphic if and only their coordinate rings $k[\mathbf{V}]$ and $k[\mathbf{W}]$ are isomorphic over k.

**Proof.** Only the "if" part is nontrivial. Given an isomorphism of coordinate rings we must produce a polynomial function from $\mathbf{V}$ to $\mathbf{W}$ which is an isomorphism. Theorem 10.13.3 does that.

**10.13.6. Example.** To show that the graph of a polynomial function $f(X,Y)$, namely $V(Z - f(X,Y)) \subset k^3$, is isomorphic to $k^2$.

**Solution.** Consider the bijections

$$\mathbf{V} = V(Z - f(X,Y)) \underset{\sigma}{\overset{\pi}{\rightleftarrows}} \mathbf{W} = V(Z),$$

where

$$\pi(x,y,z) = (x,y,0)$$
$$\sigma(x,y,0) = (x,y,f(x,y)).$$

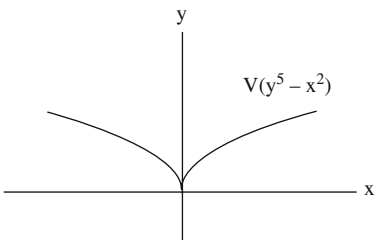



**Figure 10.19.** A curve homeomorphic but not isomorphic to **R**.

It is easy to check that their pullback maps

$$\pi^*\colon\ k[\mathbf{W}] \to k[\mathbf{V}] \quad \text{and} \quad \sigma^*\colon\ k[\mathbf{V}] \to k[\mathbf{W}]$$

are inverses of each other. The map $\sigma^*$ basically sends every polynomial F(X,Y,Z) on **V** into the polynomial where Z has been replaced by f(X,Y).

**10.13.7. Example.** The variety $V(Y^5 - X^2) \cup \mathbf{R}^2$ is homeomorphic but not isomorphic to **R**. See Figure 10.19. The projection to the x-axis clearly defines a homeomorphism between it and **R**. The proof that it is not isomorphic to **R** is somewhat involved. See [CoLO97].

Example 10.13.7 shows that the algebraic concept of isomorphism is stronger than the topological concept of homeomorphism.

After this overview of polynomial functions we are ready to tackle rational functions. Intuitively, the reader should think of a rational function on a variety as a function which can be expressed as a quotient p/q of polynomials p and q. Unfortunately, this intuitive definition runs into lots of technical problems which would require a lengthy discussion to overcome. For example, such a "function" is not defined at points where the denominator q vanishes. This in turn would make it tricky to define the composition of such functions. There are alternate, more abstract ways to define rational functions that avoid these difficulties and lead to the basic theorems more quickly. In the interest of saving time, we shall take one of these approaches so that we can give rigorous statements of theorems, even though the author normally prefers to go from the concrete to the abstract rather than vice versa. We shall sketch how the abstract concepts relate to the intuitive ones as we go along. Our approach follows that of [Shaf94].

**Definition.** Let **V** be an irreducible affine variety. The *function field* or *field of rational functions* of **V**, denoted by k(**V**) is defined to be the quotient field of the coordinate ring k[**V**] and its elements are called *rational functions* on **V**.

The next proposition shows among other things that our abstract definition of rational function really is simply another way of expressing the intuitive idea of a quotient of polynomials.

**10.13.8. Proposition.** If **V** is an irreducible affine variety, then k(**V**) is a well-defined field. Furthermore,

(1) Every rational function on **V** can be represented by a quotient $p(X)/q(X)$, where $p(X), q(X) \in k[X]$ and $q(X) \notin I(\mathbf{V})$.
(2) Two such quotients $p_1(X)/q_1(X)$ and $p_2(X)/q_2(X)$ represent the same rational function on **V** if $p_1(X)q_2(X) - p_2(X)q_1(X) \in I(\mathbf{V})$.

**Proof.** The fact that $k(\mathbf{V})$ is well defined follows from the fact that $k[\mathbf{V}]$ is an integral domain by Theorem 10.13.2. The rest is an easy exercise.

**Definition.** A rational function u on **V** is said to *regular at a point* **a** in **V** and **a** is called a *regular point* of u if it can be represented as a quotient of polynomial functions $p(X)/q(X)$, where $q(\mathbf{a}) \neq 0$. In that case we call $p(\mathbf{a})/q(\mathbf{a})$ the *value* of u at **a** and denote it by $u(\mathbf{a})$.

It follows from Proposition 10.13.8(2) that the value of a rational function at a regular point is well defined.

**10.13.9. Theorem.** Let k be an algebraically closed field. A rational function on an irreducible affine variety **V** in $k^n$ that is regular at every point of **V** is a polynomial (or regular) function.

**Proof.** See [Shaf94].

**Definition.** The set of regular points of a rational function is called its *domain* of definition.

**10.13.10. Proposition.** Let **V** be an irreducible affine variety.

(1) The domain of definition of a rational function on **V** is a nonempty open subset.
(2) A rational function on **V** is completely specified by its values on any nonempty open subset of its domain of definition.
(3) The intersection of the domain of definitions of a finite number of rational functions on **V** is again a nonempty open subset of **V**.

**Proof.** See [Shaf94].

**Definition.** Let $\mathbf{V} \subseteq k^n$ and $\mathbf{W} \subseteq k^m$ be affine varieties with **V** irreducible. A *rational function from **V** to **W**,*

$$u: \mathbf{V} \to \mathbf{W},$$

consists of an m-tuple $u = (u_1, u_2, \ldots, u_m)$ of rational functions $u_i \in k(\mathbf{V})$ with the property that if **a** is a regular point for all the $u_i$, then

$$u(\mathbf{a}) = (u_1(\mathbf{a}), u_2(\mathbf{a}), \ldots, u_m(\mathbf{a})) \in \mathbf{W}.$$

We call such a point **a** a *regular point* of u and $u(\mathbf{a})$ is called the *image* of **a**. The set of regular points of u is called the *domain* of definition of u and the set

$$u(\mathbf{V}) = \{u(\mathbf{a}) \mid \mathbf{a} \in \mathbf{V} \text{ and } \mathbf{a} \text{ is a regular point of } u\}$$

is called the *image* of $\mathbf{V}$ under u. The rational function u is said to be *dominant* if $u(\mathbf{V})$ is dense in $\mathbf{W}$ with respect to the Zariski topology.

It is easy to see that a rational function $u: \mathbf{V} \to \mathbf{W}$ is dominant if and only if $\mathbf{W}$ is the smallest variety in $\mathbf{W}$ containing $u(\mathbf{V})$ (Exercise 10.13.6). The next proposition is an immediate corollary of Proposition 10.13.10.

**10.13.11. Proposition.** The domain of definition of a rational function $u: \mathbf{V} \to \mathbf{W}$ between affine varieties with $\mathbf{V}$ irreducible is an open set and the rational function itself can be represented by an m-tuple

$$\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right) \tag{10.84}$$

of functions, where $p_i, q_i \in k[X]$ and the $q_i$ do not vanish on $\mathbf{V}$.

Clearly a rational function on an irreducible variety is the special case of where m is 1 and $\mathbf{W} = k$ above. This shows that it is natural to think of rational functions on a variety as functions defined by quotients of polynomials. It follows that polynomial functions are a special case of rational functions. They are of course defined on all of $\mathbf{V}$, but rational functions are strictly speaking not functions in general because they may not be defined everywhere. However, Theorem 10.5.6 implies that in the case of an irreducible hypersurface they are defined everywhere except on possibly a finite set of points. Finally, in analogy with Proposition 10.13.8(2), it is easy to see that representatives for rational functions are not unique in general.

It is worthwhile to see what all this means in the special case of hypersurfaces. Let $\mathbf{V}$ be a hypersurface defined by an irreducible polynomial f. The denominators $q_j$ in the representation (10.84) for the rational function u will then be polynomials which are not divisible by f. Furthermore, if p/q and r/s are representatives for rational functions u and v, respectively, which agree on $\mathbf{V}$ wherever they are both defined, then $ps - rq$ is divisible by f. The reason for this is that $ps - rq$ vanishes on all but a finite number of points of $\mathbf{V}$ and the result follows from Theorem 10.5.6.

Just as in the case of coordinate rings, let us see how function fields behave with respect to maps. Unfortunately, rational functions are not necessarily defined everywhere and so, although the idea is the same as in the case of coordinate rings and simple, the trick now is showing that everything is still well defined.

Let $u: \mathbf{V} \to \mathbf{W}$ be a rational function between irreducible affine varieties $\mathbf{V}$ and $\mathbf{W}$. Assume that u is represented by

$$\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right),$$

where $p_i, q_i \in k[X]$ and the $q_i$ do not vanish on $\mathbf{V}$. If $f \in k[\mathbf{W}]$, then let $f \circ u$ denote the rational function in $k(\mathbf{V})$ represented by

$$f\left(\frac{p_1}{q_1}, \ldots, \frac{p_m}{q_m}\right).$$

The map

$$\begin{aligned} k[\mathbf{W}] &\to k(\mathbf{V}) \\ u &\to f \circ u \end{aligned}$$

is a ring homomorphism over k that extends to a unique homomorphism

$$u^* : k(\mathbf{W}) \to k(\mathbf{V}).$$

**Definition.** The map $u^*$ is called the *pullback map* defined by u.

**10.13.12. Proposition.** The pullback map $u^*$ is a well-defined homomorphism over k.

**Proof.** See [Shaf94].

**10.13.13. Proposition.** Let $u : \mathbf{V} \to \mathbf{W}$ and $v : \mathbf{W} \to \mathbf{U}$ be rational functions between irreducible affine varieties $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{U}$.

(1) If u is dominant, then $u^*$ is one-to-one.
(2) If both u and v are dominant, then

$$v \circ u : \mathbf{V} \to \mathbf{U}$$

is a dominant rational function and

$$(v \circ u)^* = u^* \circ v^*.$$

**Proof.** See [Shaf94].

**Definition.** A dominant rational map $\varphi : \mathbf{V} \to \mathbf{W}$ between irreducible affine varieties $\mathbf{V}$ and $\mathbf{W}$ is said to be *birational* if $\varphi$ has a dominant rational inverse, that is, there is a dominant rational map $\psi : \mathbf{W} \to \mathbf{V}$, so that $\varphi \circ \psi$ and $\psi \circ \varphi$ are the identity maps wherever they are defined. We call $\mathbf{V}$ and $\mathbf{W}$ *birationally equivalent* if there is a birational map $\varphi : \mathbf{V} \to \mathbf{W}$. An affine variety is called *rational* if it is birationally equivalent to $k^n$ for some n.

**10.13.14. Theorem.** Two irreducible affine varieties are birationally equivalent if and only if they have isomorphic rational function fields.

**Proof.** See [Shaf94] or [CoLO97].

**10.13.15. Theorem.** Every irreducible affine variety is birationally equivalent to a hypersurface in some $k^n$.

**Proof.** See [Shaf94].

Note that Theorem 10.13.15 can be rephrased as saying that every irreducible variety "projects" to a hyperplane.

Let us again look at the special case of plane curves more closely.

**10.13.16. Example.** The variety $V(y^5 - x^2)$ described in Example 10.13.7 is birationally equivalent to **R**. See [CoLO97].

Example 10.13.16 shows that isomorphism is stronger than birational equivalence.

**10.13.17. Theorem.** Every rational transform of an affine rational curve is rational.

**Proof.** The theorem follows easily from Theorem B.11.16.

**10.13.18. Theorem.** Let **C** be an irreducible plane curve defined by an equation $f(X,Y) = 0$. The following two conditions are equivalent:

(1) There exist two rational functions $p(t), q(t) \in k(t)$ so that

  (a) $f(p(t),q(t)) = 0$ for all but a finite number of t, and
  (b) for all but a finite number of points (x,y) on **C** there is a unique $t \in k$ satisfying $x = p(t)$ and $y = q(t)$.

(2) The curve **C** is rational.

**Proof.** Because of Theorem 10.13.14, condition (2) is equivalent to

(3) The function field k(**C**) is isomorphic to the field of rational functions k(t) for a transcendental variable t.

We follow the proof in [Walk50] and show that (1) and (3) are equivalent.

(1) $\Rightarrow$ (3): Condition (1b) implies that the functions p and q are not both constants. Assume that p is not constant. It will then be transcendental over k. Since $k(p,q) \subseteq k(t)$, Theorem B.11.5 implies that $k(p,q) = k(\lambda)$ for some $\lambda \in k(t)$. But $k(\mathbf{C}) \approx k(p,q)$, and so we have proved (3).

(3) $\Rightarrow$ (1): If the function field of **C** is isomorphic to $k(\lambda)$, where $\lambda$ is transcendental over k, then **C** is birationally equivalent to the line $s = 0$ in (s,t)-space, that is,

$$x = p(t) \quad \text{and} \quad y = q(t)$$

for rational functions p and q. These functions satisfy (1a) and (1b) because, except for a finite number of points on **C** and $s = 0$, the birational equivalence is a bijection between the points of these two curves.

Theorem 10.13.18 is proved.

**10.13.19. Theorem.** If a plane curve $f(X,Y) = 0$ has the property that there exist non-constant rational functions $p(t)$, $q(t) \in k(t)$ so that $f(p(t),q(t)) = 0$ for all but a finite number of t, then the curve is rational.

**Proof.** This theorem was really already proved in the process of proving Theorem 10.13.18 since condition (1b) was **only** used to show that the $p(t)$ and $q(t)$ were not both constant.

Theorems B.11.5, 10.13.17, and 10.13.19 are all equivalent and often referred to as Lüroth's Theorem. More generally, stated in field theory terms we have:

> **The general *Lüroth problem***: If K is a subfield of $k(X_1,X_2, \ldots, X_n)$ such that $k(X_1,X_2, \ldots, X_n)$ is a finite separable extension of K, is K isomorphic to a rational function field?

The answer is "yes" when $n = 1$, even without the separability condition. If $n = 2$, the answer is "yes" if k is algebraically closed. The answer is "no" if $n \geq 3$, even if $k = \mathbf{C}$. See [Shaf94].

Sometimes condition (1) in Theorem 10.13.18 is what is used for a definition of a rational plane curve. The theorem shows that this alternate definition and ours above are equivalent. In other words, to show that a curve is rational it suffices to parameterize a curve with rational functions in a one-to-one fashion and one does not have to check anything about its inverse (whether it is rational). The finite number of exceptions in (1a) arise from denominators of rational functions vanishing **and** from singularities of the curve.

Let us take a closer look at what all this means in the special case of affine curves. Assume that **C** is an affine plane curve defined by an equation

$$f(X,Y) = 0.$$

The rational functions on **C** are function of X and Y. Furthermore, X and Y are algebraically dependent since they satisfy an algebraic equation. We can strengthen that.

**10.13.20. Lemma.** Any two elements of $k(\mathbf{C})$ are algebraically dependent over k.

**Proof.** See [Walk50].

**10.13.21. Lemma.** $k(\mathbf{C})$ has transcendence degree 1 over k.

**Proof.** First of all, $k(\mathbf{C})$ has transcendence degree at least 1 since the "functions" X or Y are not algebraic over k. Lemma 10.13.20 finishes the proof.

Looking ahead to Section 10.16 where we define the dimension of a variety, we shall see that it is Lemma 10.13.21 that establishes the fact that curves have the dimension we expect them to have, namely, they are one-dimensional spaces (Theorem 10.16.9).

**10.13.22. Theorem.** A field K over an algebraically closed field of characteristic 0 is the function field of an irreducible plane curve if and only if

(1) $\mathrm{tr}_k(K) = 1$.
(2) K has a finite basis over k.

**Proof.** See [Walk50].

Next, we describe another way of looking at rational functions.

**Definition.** Let K be an extension field of the complex numbers. Let $f \in \mathbf{C}[X,Y]$ and assume that $f(X,Y) = 0$ is the minimal equation for a curve $\mathbf{C} \subseteq K^2$. A point $(a,b) \in K^2$ is called a *generic point* for $\mathbf{C}$ if $g \in \mathbf{C}[X,Y]$ and $g(a,b) = 0$ implies that $g(x,y) = 0$ for all points $(x,y) \in K^2$ on the curve $\mathbf{C}$.

The parameterizations discussed in the last section provide the context we have in mind for generic points. It follows that if (a,b) is a generic point for the curve $\mathbf{C}$, then $g(a,b) = 0$ implies that f divides g. Also, the next theorem shows that generic points are only meaningful in the case of irreducible curves.

**10.13.23. Theorem.** Let $\mathbf{C}$ be a curve as in the definition of a generic point.

(1) If $\mathbf{C}$ is reducible, then it does not have any generic points.
(2) If $\mathbf{C}$ is irreducible, then every point $(a,b) \in \mathbf{C}$ with not both a and b complex numbers is a generic point. In particular, $\mathbf{C}$ has an infinite number of generic points.

**Proof.** See [Seid68].

**Definition.** Let K be an extension field of k. Two points $(x_1,y_1)$ and $(x_2,y_2)$ in $K^2$ are said to be *isomorphic* is there is an isomorphism $\sigma: k[x_1,y_1] \to k[x_2,y_2]$ over k that sends $x_1$ to $x_2$ and $y_1$ to $y_2$.

**10.13.24. Theorem.** Any two generic points $(x_1,y_1)$ and $(x_2,y_2)$ for a curve $\mathbf{C}$ as above are isomorphic.

**Proof.** See [Seid68].

One can now show ([Seid68]) that if (x,y) is a generic point for an irreducible curve, then an alternate definition for the field of rational functions on the curve is to say that it is the field $\mathbf{C}(x,y)$. That field is well defined, up to isomorphism over the complex numbers $\mathbf{C}$, by Theorem 10.13.24. The field $\mathbf{C}(x,y)$ will be a finite extension of $\mathbf{C}$ with transcendency degree 1. This fits in with Theorem 10.13.22.

Here are some more useful observations about birational maps on affine varieties.

**10.13.25. Proposition.** If $\varphi: \mathbf{C} \to \mathbf{D}$ is a birational map between irreducible affine plane curves $\mathbf{C}$ and $\mathbf{D}$, then $\varphi^{-1}(\mathbf{q})$ is finite for every $\mathbf{q} \in \mathbf{D}$.

**Proof.** This is an easy consequence of Theorem 10.5.6. Let $\psi: \mathbf{D} \to \mathbf{C}$ be the inverse of $\varphi$. Let $\mathbf{U} \subset \mathbf{C}$ be the set of points at which $\varphi$ is defined. Let $\mathbf{V} \subset \mathbf{D}$ be the set of points at which $\psi$ is defined. Both $\mathbf{C} - \mathbf{U}$ and $\mathbf{D} - \mathbf{V}$ are finite sets. It follows that the complement of $\varphi^{-1}(\mathbf{V}) \cap \mathbf{U}$ in $\mathbf{C}$ and the complement of $\psi^{-1}(\mathbf{U}) \cap \mathbf{V}$ in $\mathbf{D}$ is finite and $\varphi$ is a bijection between $\varphi^{-1}(\mathbf{V}) \cap \mathbf{U}$ and $\psi^{-1}(\mathbf{U}) \cap \mathbf{V}$.

Another interesting fact about birational equivalences is that they define a bijection between places and generic points. See [Seid68].

**10.13.26. Theorem.** Any irreducible affine plane curve in $\mathbf{C}^2$ is birationally equivalent to a plane curve with only ordinary singularities.

**Proof.** This is only a restatement of Lemma 10.12.15 since the quadratic transformations used there are birational equivalences. See [Walk50].

If we want to end up with a plane curve, then Theorem 10.13.26 cannot be improved. We cannot prevent ordinary singularities, that is, points where the curve crosses itself transversally. On the other hand, if we allow ourselves to move to curves in higher dimensions, then one can show the following:

**10.13.27. Theorem.** Any affine variety in $\mathbf{C}^2$ is birationally equivalent to a variety with no singularities.

**Proof.** See [Harr92] or [Shaf94].

Theorem 10.13.27 does not say anything about the dimension of the space that contains the variety with no singularities. One can show that any plane curve in $\mathbf{P}^2(\mathbf{C})$ is birationally equivalent to a curve in $\mathbf{P}^3(\mathbf{C})$ that has no singularities ([BriK81]). See also Theorem 10.14.7 in the next section.

Next, we discuss a class of functions defined on varieties that are especially interesting, namely, the "finite" functions. They give an algebraic characterization of covering spaces. First, we need a definition.

**Definition.** Let S be a subring of a commutative ring R with identity and assume that S contains that identity. An element $r \in R$ is said to be *integral over S* if

$$r^m + s_{m-1}r^{m-1} + \ldots s_1 r + s_0 = 0 \quad \text{for some } s_i \in S.$$

We say that R is *integral over S* if every element of R is integral over S.

Let $f: \mathbf{V} \to \mathbf{W}$ be a dominant regular function between affine varieties. We know from Proposition 10.13.4 that $f^*: k[\mathbf{W}] \to k[\mathbf{V}]$ is one-to-one and so we may consider $k[\mathbf{W}]$ as a subring of $k[\mathbf{V}]$.

**Definition.** We say that f is a *finite map* if $k[\mathbf{V}]$ is integral over $k[\mathbf{W}]$.

**10.13.28. Theorem.**

(1) The inverse image of every point for a finite map is a finite set.
(2) If k is algebraically closed, then a finite map is onto and takes closed sets to closed sets.

**Proof.** See [Shaf94].

To be a finite map is a local property.

**10.13.29. Theorem.** Let $f: \mathbf{V} \to \mathbf{W}$ be a regular map between affine varieties. If every point of $\mathbf{W}$ has a neighborhood so that the inverse image under f of every point in that neighborhood is a finite set, then f is finite.

**Proof.** See [Shaf94].

Up to now we have discussed functions defined on affine varieties. It is possible to define regular and rational functions for projective varieties, but things get more complicated. Recall how we restricted ourselves to homogeneous polynomials in order to get a well-defined definition for a projective variety. We need to make similar restrictions here.

**Definition.** A *rational homogeneous polynomial* in n + 1 variables $X_1, X_2, \ldots, X_{n+1}$ is a rational polynomial function of the form

$$h(X_1, X_2, \ldots, X_{n+1}) = \frac{f(X_1, X_2, \ldots, X_{n+1})}{g(X_1, X_2, \ldots, X_{n+1})},$$

where f and g are homogeneous polynomials of the same degree. Let $\mathbf{V} = V(g)$ be the set of points where the denominator g vanishes. The function

$$\begin{array}{ccc} \mathbf{P}^n(k) - \mathbf{V} & \to & k \\ [X_1, X_2, \ldots, X_{n+1}] & \to & h(X_1, X_2, \ldots, X_{n+1}) \end{array}$$

is called a *rational function* on $\mathbf{P}^n(k)$ and will also be denoted by h or f/g.

It is easy to check that a rational homogeneous polynomial in n + 1 variables defines a well-defined rational function on $\mathbf{P}^n(k)$.

**Definition.** Let h = f/g be a rational homogeneous polynomial in n + 1 variables. If $\mathbf{p} \in \mathbf{P}^n(k)$ and $g(\mathbf{p}) \neq 0$, then the rational function h on $\mathbf{P}^n(k)$ is said to be *regular at* $\boldsymbol{p}$. If h is regular at all points of a set $\mathbf{X} \subseteq \mathbf{P}^n(k)$, then it is called a *regular function* on $\mathbf{X}$. The set of regular functions on $\mathbf{X}$ is denoted by $k[\mathbf{X}]$.

Clearly, the natural addition and multiplication make $k[\mathbf{X}]$ into a ring.

**10.13.30. Proposition.** $k[\mathbf{X}] \approx \dfrac{k[X_1, X_2, \ldots, X_{n+1}]}{I(\mathbf{X})}$.

**Proof.** See [Harr92].

It is also easy to show that if we restrict a regular function on a projective variety to some affine part with respect to some parameterization, then the new notion of regular agrees with the same notion for affine varieties. There is one big difference however. The only regular functions on irreducible projective varieties are the constants (Theorem 10.13.32). This does not happen in the affine case.

**Definition.** Let $f: \mathbf{V} \to \mathbf{W}$ be a map between projective varieties with $\mathbf{W} \subseteq \mathbf{P}^n(k)$. Let $\mathbf{p} \in \mathbf{V}$. Assume that the following two properties hold:

(1) One can choose homogeneous coordinates for $\mathbf{P}^n(k)$ so that for some neighborhood $\mathbf{U}$ of $\mathbf{p}$ in $\mathbf{V}$ the image $f(\mathbf{U})$ belongs to the affine part of $\mathbf{P}^n(k)$ with respect to these coordinates. With this choice we can (and will) consider the map $g = f|\mathbf{U}$ as a map from $\mathbf{U}$ to $k^n$.
(2) If $g(\mathbf{q}) = (g_1(\mathbf{q}), g_2(\mathbf{q}), \ldots, g_n(\mathbf{q}))$, then the $g_i : \mathbf{U} \to k$ are rational functions that are regular at $\mathbf{p}$.

In this case, we say that the function f is *regular at* $\boldsymbol{p}$. The function f is said to be *regular on* $\boldsymbol{V}$ if f is regular at all points of $\mathbf{V}$.

Again one can show that the notion of being regular, or regular at a point, does not depend on n, $\mathbf{U}$, or the homogeneous coordinates that are chosen. As we warned earlier, defining regularity of maps between projective varieties is complicated. It involves checking that for each point one can find a coordinate system with respect to which the function is affine-valued and can be expressed in terms of regular rational functions. One type of function that it is easily seen to be a regular map is one that can be expressed globally as an $(n + 1)$-tuple of homogeneous polynomials of the same degree that do not have any common zeros. It is unfortunate that, as the next example shows, not all regular maps can be obtained in this way, because that would have allowed for a much simpler definition.

**10.13.31. Example.** Consider the variety $\mathbf{V}$ in $\mathbf{P}^2(\mathbf{C})$ defined by

$$X^2 + Y^2 - Z^2 = 0$$

and the "stereographic projection"

$$f: \mathbf{V} \to \mathbf{P}^1(\mathbf{C})$$
$$f([X,Y,Z]) = [X, Z - Y].$$

The polynomials X and $Z - Y$ have a common zero at (0,1,1), so that f is not defined at [0,1,1], but we define it there by setting f([0,1,1]) equal to [1,0]. We claim that f is a regular map. Furthermore, it cannot be expressed as a pair of homogeneous polynomials without common zeros.

**Proof.** Consider the open cover $\{\mathbf{O}_1, \mathbf{O}_2\}$ of $\mathbf{P}^1(\mathbf{C})$, where

$$\mathbf{O}_1 = \{[x,y] \mid x \neq 0\} \quad \text{and} \quad \mathbf{O}_2 = \{[x,y] \mid y \neq 0\}.$$

Let

$$\mathbf{U}_1 = f^{-1}(\mathbf{O}_1) = \mathbf{P}^2(\mathbf{C}) - \{[0,1,-1]\} \quad \text{and} \quad \mathbf{U}_2 = f^{-1}(\mathbf{O}_2) = \mathbf{P}^2(\mathbf{C}) - \{[0,1,1]\}.$$

Then f is regular on $\mathbf{U}_2$, because f sends [X,Y,Z] to the point whose affine coordinate is $X/(Z - Y)$. On $\mathbf{U}_1$ the point [X,Y,Z] gets sent to the point whose affine coordinate is $(Z - Y)/X$. It may seem as if there is a problem with regularity at [0,1,1], but fortunately we can rewrite the quotient as

$$\frac{Z-Y}{X}=\frac{Z^2-Y^2}{X(Z+Y)}=\frac{X^2}{X(Z+Y)}=\frac{X}{Z+Y},$$

which is well defined at [0,1,1]. This shows that f is regular everywhere. The rest is left as Exercise 10.13.8.

It is the fact that we were forced to figure out a way to rewrite the formula for the function f in Example 10.13.31 over the open set $\mathbf{U}_1$ that shows that determining whether or not a function is regular would not be any simpler if we were to define regular maps in terms of tuples of homogeneous polynomials where common zeros are allowed.

Like in the affine case, a regular map

$$f:\ \mathbf{V}\to\mathbf{W}$$

defines a homomorphism

$$f^*:\ k[\mathbf{W}]\to k[\mathbf{V}].$$

**Definition.** A regular map $f:\mathbf{V}\to\mathbf{W}$ between two projective varieties is called an *isomorphism* if it has an inverse that is also a regular map.

Next, we define rational functions between projective varieties. We cannot simply define a rational function on a projective variety to be a regular function, because of the following result:

**10.13.32. Theorem.** A regular function defined on an irreducible projective variety is constant.

**Proof.** See [Shaf94].

**Definition.** Let $\mathbf{V}$ be a projective variety in $\mathbf{P}^n(k)$. A *rational function* on $\mathbf{V}$ is a rational homogeneous polynomial in $k(X_1,X_2,\ldots,X_{n+1})$. The set of rational functions on $\mathbf{V}$ is called the *function field* of $\mathbf{V}$ and is denoted by $k(\mathbf{V})$. A rational function h is *regular at a point* $\mathbf{p}\in\mathbf{V}$, if it can be written in the form $h = f/g$, where f and g are homogeneous polynomials of the same degree and $g(\mathbf{p})\neq 0$. In that case, $h(\mathbf{p}) = f(\mathbf{p})/g(\mathbf{p})$ is called the *value* of h at $\mathbf{p}$. The set of regular points of h is called the *domain* of h.

The function field of a projective variety is clearly a field. In fact, it is easy to see that

$$k(\mathbf{V})\approx k(Y_1,Y_2,\ldots,Y_n),$$

where $Y_1, Y_2, \ldots, Y_n$ are indeterminates, because

$$\frac{f(X_1,X_2,\ldots X_{n+1})}{g(X_1,X_2,\ldots X_{n+1})}=\frac{f\left(\frac{X_1}{X_{n+1}},\frac{X_2}{X_{n+1}},\ldots\frac{X_n}{X_{n+1}},1\right)}{g\left(\frac{X_1}{X_{n+1}},\frac{X_2}{X_{n+1}},\ldots\frac{X_n}{X_{n+1}},1\right)}=\frac{f(Y_1,Y_2,\ldots Y_n,1)}{g(Y_1,Y_2,\ldots Y_n,1)},$$

where $Y_i = X_i/X_{n+1}$ are considered indeterminates.

**Definition.** Let $\mathbf{V}$ be a projective variety in $\mathbf{P}^n(k)$. A *rational map* $f: \mathbf{V} \to \mathbf{P}^m(k)$ is a map of the form

$$\mathbf{p} \to [f_1(\mathbf{p}), f_2(\mathbf{p}), \ldots, f_{m+1}(\mathbf{p})]$$

where the $f_i$ are homogeneous polynomials in $k[X_1, X_2, \ldots, X_{n+1}]$ of the same degree and at least one of the $f_i$ must be nonzero at every point $\mathbf{p} \in \mathbf{V}$. The map f will be denoted by the tuple $(f_1, f_2, \ldots, f_{m+1})$

Clearly, two rational maps

$$f = (f_1, f_2, \ldots f_{m+1}),\ g = (g_1, g_2, \ldots, g_{m+1}):\ \mathbf{V} \to \mathbf{P}^m(k)$$

are equal if and only if $f_i g_j = g_i f_j$ on $\mathbf{V}$ for all i and j. Since, given a rational map $f = (f_1, f_2, \ldots, f_{m+1})$, we could divide through by one of the $f_i$, we see that a rational map is defined by m + 1 rational functions on $\mathbf{V}$. Projections, defined below, are good examples of regular rational maps.

**Definition.** Assume that $\mathbf{X}$ is a d-dimensional linear subspace of $\mathbf{P}^n(k)$ defined the (n – d) equations

$$L_1 = L_2 = \ldots = L_{n-d} = 0,$$

where the $L_i$ are linear homogeneous polynomials. Define

$$\pi_{\mathbf{X}}:\ \mathbf{P}^n(k) \to \mathbf{P}^{n-d-1}(k)$$

by

$$\pi_{\mathbf{X}}(\mathbf{p}) = [L_1(\mathbf{p}), L_2(\mathbf{p}), \ldots, L_{n-d}(\mathbf{p})].$$

If $\mathbf{V} \subseteq \mathbf{P}^n(k)$, then $\pi_{\mathbf{V}} = \pi_{\mathbf{X}}|\mathbf{V}$ is called the *projection of* $\boldsymbol{V}$ *with center* $\boldsymbol{X}$.

The map $\pi_{\mathbf{X}}$ is clearly a regular map on $\mathbf{P}^n(k) - \mathbf{X}$. In fact, if $\mathbf{V}$ is any projective variety that is disjoint from $\mathbf{X}$, then $\pi_{\mathbf{X}}|\mathbf{V}$ is a regular rational map. To get a feel for what the map $\pi_{\mathbf{X}}$ does, let $\mathbf{Y}$ be any (n – d – 1)-dimensional linear subspace of $\mathbf{P}^n(k)$. Then $\pi_{\mathbf{X}}$ maps $\mathbf{p} \in \mathbf{P}^n(k)$ to the (unique) intersection of the linear subspace of $\mathbf{P}^n(k)$ generated by $\mathbf{p}$ and $\mathbf{X}$ with $\mathbf{Y}$. See Figure 10.20 for the case where d = 0.

Here is another example of a regular rational map.

**Definition.** Fix n and d and define

$$v_d:\ \mathbf{P}^n(k) \to \mathbf{P}^N(k)$$

by the condition

$$v_d[x_1, x_2, \ldots, x_{n+1}] = [\ \ldots, m_I, \ldots\ ],$$

where $m_I$ ranges over all monomials of degree d in $x_1, x_2, \ldots, x_{n+1}$, of which there are $N = \binom{n+d}{d} - 1$. The map $v_d$ is called the *Veronese imbedding* of $\mathbf{P}^n(k)$ in $\mathbf{P}^N(k)$ and its image $v_d(\mathbf{P}^n(k))$ is called the *Veronese variety*.

For example, if n = 2, then

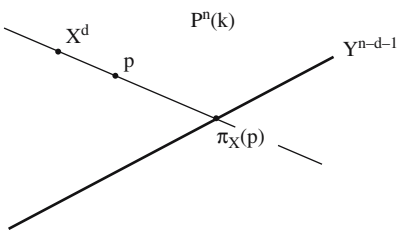



**Figure 10.20.** The projection of $P^n(k)$ with center X.

$$v_2 : \mathbf{P}^2(k) \to \mathbf{P}^5(k)$$

is defined by

$$v_2[x_1, x_2, x_3] = [x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3].$$

It is easy to see that the Veronese variety is in fact a projective variety in $\mathbf{P}^N(k)$. Sometimes any variety isomorphic to it is also called a Veronese variety.

**Definition.** Let $\mathbf{V}$ and $\mathbf{W}$ be projective varieties and assume that $\mathbf{W} \subseteq \mathbf{P}^m(k)$. A *rational map*

$$f : \mathbf{V} \to \mathbf{W}$$

is a rational map

$$f : \mathbf{V} \to \mathbf{P}^m(k)$$

with the property that $f(\mathbf{V}) \subseteq \mathbf{W}$. The map f is called *birational* and we say that $\mathbf{V}$ and $\mathbf{W}$ are *birationally equivalent* if f has an inverse $g : \mathbf{W} \to \mathbf{V}$ that is a rational map.

**10.13.33. Theorem.** The image of a projective variety under a rational map is a projective variety.

**Proof.** See [Shaf94].

**10.13.34. Theorem.** Any two nonsingular projective curves that are birationally equivalent are isomorphic.

**Proof.** See [Shaf94].

**10.13.35. Corollary.** Any two nonsingular projective curves that are birationally equivalent are homeomorphic.

We generalize the notion of a finite map to projective varieties by making it a local property.

**Definition.** A regular map $f: \mathbf{V} \to \mathbf{W}$ between projective varieties is called a *finite map* if every point $\mathbf{q}$ in $\mathbf{W}$ has an affine neighborhood $\mathbf{B_q}$ in $\mathbf{W}$, so that $\mathbf{A_q} = f^{-1}(\mathbf{B_q})$ is an affine set in $\mathbf{V}$ and

$$f \mid \mathbf{A_q} : \mathbf{A_q} \to \mathbf{B_q}$$

is a finite map between affine varieties.

Projections are important examples of finite maps.

**10.13.36. Theorem.** Let $\mathbf{V}$ be any projective variety in $\mathbf{P}^n(k)$ that is disjoint from a d-dimensional linear subspace $\mathbf{X}$ in $\mathbf{P}^n(k)$. Then the projection of $\mathbf{V}$

$$\pi_{\mathbf{V}} : \mathbf{V} \to \mathbf{P}^{n-d-1}(k)$$

with center $\mathbf{X}$ defines a finite map

$$\mathbf{V} \to \pi_{\mathbf{V}}(\mathbf{V}).$$

**Proof.** See [Shaf94].

A very useful application of Theorem 10.13.36 is the following generalization needed later.

**10.13.37. Theorem.** Let $p_1, p_2, \ldots, p_{s+1} \in k[X_1, X_2, \ldots, X_{n+1}]$ be homogeneous polynomials of degree d. If the $p_i$ have no common zeros on a projective variety $\mathbf{V} \subseteq \mathbf{P}^n(k)$, then

$$f([\mathbf{x}]) = [p_1(\mathbf{x}), p_2(\mathbf{x}), \ldots, p_{s+1}(\mathbf{x})] \in \mathbf{P}^s(k)$$

defines a finite map $f: \mathbf{V} \to f(\mathbf{V})$.

**Proof.** Consider the Veronese imbedding

$$v_d : \mathbf{P}^n(k) \to \mathbf{P}^N(k).$$

Now, if

$$p(X_1, X_2, \ldots, X_{n+1}) = \sum_{i_1 + \ldots + i_{n+1} = d} a_{i_1 i_2 \ldots i_{n+1}} X_1^{i_1} X_2^{i_2} \ldots X_{n+1}^{i_{n+1}}$$

is a homogeneous polynomial of degreed, then let $\mathbf{L}_p$ be the hyperplane of $\mathbf{P}^N(k)$ defined by the linear equations

$$\sum_{i_1 + \ldots + i_{n+1} = d} a_{i_1 i_2 \ldots i_{n+1}} X_{i_1 i_2 \ldots i_{n+1}} = 0,$$

where $X_{i_1 i_2 \ldots i_{n+1}}$ is the indicated indexed variable in the collection, $X_1, X_2, \ldots, X_N$. The property of the Veronese imbedding that we want to use here is that $v_d(V(p)) = v_d(\mathbf{P}^n(k)) \cap \mathbf{L}_p$. To prove our theorem, let $\pi$ be the projection of $\mathbf{P}^N(k)$ defined by the hyperplanes $\mathbf{L}_{p_i}$. One can show that $f = \pi \circ v_d$, so that our theorem now follows from Theorem 10.13.36.

**10.13.38. Theorem.** (The Noether Normalization Theorem)

(1) Any irreducible projective variety $\mathbf{V}$ in $\mathbf{P}^n(k)$ admits a finite map $f: \mathbf{V} \to \mathbf{P}^m(k)$ for some $m \le n$.
(2) Any irreducible affine variety $\mathbf{V}$ in $k^n$ admits a finite map $f: \mathbf{V} \to k^m$ for some $m \le n$.

**Proof.** See [Shaf94]. We sketch the proof of (1). Assume that $\mathbf{V} \neq \mathbf{P}^n(k)$ and let $\mathbf{p} \in \mathbf{P}^n(k) - \mathbf{V}$. The projection $\varphi$ of $\mathbf{V}$ with center $\mathbf{p}$ will be regular. The image $\varphi(\mathbf{V})$ in $\mathbf{P}^{n-1}(k)$ will be a projective variety and $\varphi: \mathbf{V} \to \varphi(\mathbf{V})$ is a finite map by Theorem 10.13.36. If $\varphi(\mathbf{V}) \neq \mathbf{P}^{n-1}(k)$ then we can repeat this process. Since the composite of a finite number of finite maps is finite we will finally get our map f.

Our coordinate rings used functions (polynomials) that were defined on the whole variety. However, one can also get useful information from local properties. Instead of globally defined functions one can look at *local rings*. These are defined for every point and are rings of functions that are only defined in a neighborhood of the point. See [Shaf94]. They also show up in complex analysis.

Finally, we want to draw the reader's attention to a property of the rational parameterization, call it $\varphi(t)$, of a conic that we described in the discussion after Example 10.13.1 and its bearing on the following type of problem: Given a subfield k of a field K and a curve in $K^2$ defined by $f(X,Y) = 0$, find all the points of the curve with coordinates in k. For example, we might want the rational points of the curve in Example 10.13.1. In the case of a conic, if all the coefficients of $f(X,Y)$ and the coordinates of the given point $\mathbf{p}_0$ of the curve belonged to k, then our $\varphi(t)$ will generate the desired points as t ranges over k. Other curves admit similar parameterizations.

## 10.14 Space Curves

In addition to plane curves, another class of spaces that have great practical interest are space curves, specifically curves in $\mathbf{R}^3$. The simplest and intuitive definition is:

**Definition.** A *curve* or *algebraic curve* is an irreducible algebraic variety of dimension 1.

The only problem with this definition is that we have not yet defined what the dimension of an algebraic variety is. We shall do so in Section 10.16, but the concept of dimension and the associated topic of higher dimensional varieties, although very important in algebraic geometry, is too advanced for us to do anything more than give a brief overview. For that reason, to avoid a lengthy digression at this point, we shall give an equivalent, but ad hoc, definition of a curve that does not use dimension and yet will enable us to study some of their properties. Our approach, which follows that given in [Walk50], will seem rather roundabout and a kind of "trick." At the end of this section we shall rephrase the definition in terms of ideals. There is one property of dimension that the reader should be aware of right now though, otherwise it might be puzzling why we continually restrict ourselves to transcendence degree one in the discussion that follows. In the case of an irreducible variety, its dimension is the same as the transcendence degree of its function field (see Theorem 10.16.9).

Note that the function field K of an affine curve in the "plane" $k^2$ can be expressed in the form $K = k(x,y)$, where x is transcendental over k and y is algebraic over $k(x)$.

The definition of a space curve will be based on parameterizations and is gotten by expressing K as an extension of k by n elements, $n > 2$ . The extra elements in the basis will let us interpret the curve as a curve in $k^n$. This new way of looking at the plane curve will tell us what the definition of a curve in $k^n$ should be.

After these preliminary comments, assume that

$$K = k(\xi_1, \xi_2, \ldots, \xi_n)$$

is an arbitrary extension of k with transcendence degree 1. By Theorems B.11.4 and B.11.8, K will have the form k(x,y) described in the previous paragraph and correspond to an affine plane curve $\mathbf{C}_f$ defined by an equation

$$f(X, Y) = 0.$$

Assume that $f(X,Y) \neq aX + b$ , so that $y \neq 0$, which is the interesting case. Assume further that the projective curve that it represents is irreducible. It follows that we can express the $\xi_i$ in terms of x and y and conversely that we can express x and y in terms of the $\xi_i$. In other words, there exist $g, h \in k(X_1, X_2, \ldots, X_n)$ and $z_i \in k(X,Y)$ so that

$$\xi_i = z_i(x, y)$$

and

$$x = g(\xi_1, \xi_2, \ldots, \xi_n), \quad y = h(\xi_1, \xi_2, \ldots, \xi_n).$$

Any parameterization (x(t),y(t)) representing a place $\Pi$ of $\mathbf{C}_f$ induces a "parameterization"

$$(z_1(x(t), y(t)), z_2(x(t), y(t)), \ldots, z_n(x(t), y(t))) = (x_1(t), x_2(t), \ldots, x_n(t)).$$

Switching to projective coordinates we get a "parameterization"

$$(z_1(t), z_2(t), \ldots, z_{n+1}(t)), \quad \text{where } x_i = z_i / z_{n+1}.$$

Choose our representative for the projective point so that $\min(\mathrm{ord}(z_i(t))) = 0$. Let

$$\mathbf{q} = (z_1(0), z_2(0), \ldots, z_{n+1}(0)).$$

Let $\mathbf{C}_P$ be the subset of $\mathbf{P}^n(k)$ consisting of all these points $\mathbf{q}$ as we let $\Pi$ range over all places of $\mathbf{C}_f$. Let $\mathbf{C}$ be the affine subset of $k^n$ associated to $\mathbf{C}_P$.

**Definition.** The subset $\mathbf{C}_P$ in $\mathbf{P}^n(k)$ that is obtained in this way is called an irreducible *projective space curve*. The subset $\mathbf{C}$ in $k^n$ is called an irreducible *affine space curve*.

Equivalence classes of parameterizations of $\mathbf{C}_P$ or $\mathbf{C}$ are called *places* for the curves. All the terminology that we developed for places for plane curves apply here. One can also define birational correspondences

$$u: \ K = k(\theta_1, \theta_2, \ldots, \theta_n) \to K = k(\theta_1', \theta_2', \ldots, \theta_n'),$$

which have the form

$$\theta'_j = \Psi_j(\theta_1, \theta_2, \ldots, \theta_n)$$

in the affine case. We get a unique curve for each basis of K.

All of this may sound good, but we still need to show that $\mathbf{C}_P$ and $\mathbf{C}$ are not simply sets but varieties, specifically, curves. Various facts have to be established first before one can prove the following:

**10.14.1. Theorem.** If $\mathbf{C}$ is a variety, then $\mathbf{C}$ is a space curve if and only if $\mathbf{C}$ is an irreducible curve.

**Proof.** The theorem is a consequence of the next two theorems.

We relate the space curve $\mathbf{C}$ in $k^n$ to an ideal. Define

$$I = \{f \in k[X_1, X_2, \ldots, X_n] \mid f(\xi_1, \xi_2, \ldots, \xi_n) = 0\}.$$

**10.14.2. Theorem.** The ideal I in $k[X_1, X_2, \ldots, X_n]$ is a prime ideal and K is isomorphic to the quotient field of $k[X_1, X_2, \ldots, X_n]/I$.

**Proof.** It is trivial to check that I is an ideal. The rest follows easily by analyzing the map

$$\begin{array}{ccc} k[X_1, X_2, \ldots, X_n] & \to & k[\xi_1, \xi_2, \ldots, \xi_n]. \\ X_i & \to & \xi_i \end{array}$$

This map is onto and has kernel I.

Theorem 10.14.2 shows that I determines the curve $\mathbf{C}$ completely. One also has

**10.14.3. Theorem.**

(1) $f \in I$ if and only if $f(\mathbf{p}) = 0$ for all $\mathbf{p} \in \mathbf{C}$.
(2) Let $\mathbf{p} \in k^n$ . If $f(\mathbf{p}) = 0$ for all $f \in I$, then $\mathbf{p} \in \mathbf{C}$.

**Proof.** See [Walk50].

In the above we started with a curve and got an ideal that defined it. We can go the other way and get a curve starting with an ideal.

**10.14.4. Theorem.** Let k be an algebraically closed field of characteristic 0. An ideal I in $k[X_1, X_2, \ldots, X_n]$ is the ideal associated to an irreducible space curve if and only if I is prime and its transcendence degree over k, that is, $\mathrm{tr}_k(k(I))$, is equal to 1.

**Proof.** See [Walk50].

It follows from Theorem 10.14.3 that $\mathbf{C} = V(I(\mathbf{C}))$. This fact and Theorem 10.14.4 show that we could have defined a space curve as being the set of zeros of certain ideals. This algebraic approach to the definition of a space curve would certainly be much cleaner than our messy construction for the points of such a curve.

**10.14.5. Example.** Consider the space curve called the *twisted cubic*

$$\mathbf{V} = V(y - x^2, z - x^3) \subset \mathbf{R}^3.$$

Here are some facts about it:

(1) It can be parameterized by the map $t \to (t,t^2,t^3)$, so that it is the same curve as the one defined in Exercise 9.4.2
(2) $I(\mathbf{V}) = \langle y - x^2, z - x^3 \rangle$ (See [CoLO97])
(3) $\mathbf{V}$ is irreducible (Theorem 10.8.14).
(4) The curve is an example that shows that homogenizing its equation does not lead to the smallest projective variety containing it. See Exercise 10.3.4. On the other hand, a Gröbner basis for it is

$$G = \{x^2 - y, xy - z, xz - y^2\},$$

and a homogenization of this basis does lead to a basis

$$\{x^2 - wy, xy - wz, xz - y^2\}$$

for the projective ideal for that smallest projective variety. See [CoLO97].
(5) Property (3), Theorem 10.8.12, and Theorem 10.14.4 imply that the twisted cubic is a space curve.

The argument that showed that the twisted cubic is irreducible can be extended to prove the following:

**10.14.6. Theorem.** If k is an infinite field, then any affine variety $\mathbf{V}$ in $k^n$ that has a rational parameterization is irreducible.

**Proof.** See [CoLO97]. Theorem 10.8.14 was a special case of this theorem.

Applying Theorem 10.14.6 in the special case of polynomial parameterizations gives lots of examples of irreducible varieties, so that, using Theorem 10.8.12 and Theorem 10.14.4, we get lots of examples of space curves.

**10.14.7. Theorem.** Every space curve is birationally equivalent to a plane curve that is either nonsingular or has only ordinary singularities.

**Proof.** See [Walk50] or [Seid68].

The fact that every space curve is birationally equivalent to a plane curve was already suggested from Theorems 10.13.14 and 10.13.15. The only problem is that they applied to hypersurfaces, which (nonplane) curves are not. Now, Theorem 10.13.27 showed that every space curve is birationally equivalent to a nonsingular curve; however, this curve may **not** be planar. There are nonsingular space curves that are not birationally equivalent to a nonsingular plane curve. We can reduce any singularities to ordinary double points, however.

Two more facts about space curves are:

**10.14.8. Theorem.** If k is an algebraically closed field, then any space curve in $k^3$ is contained in an algebraic surface.

**Proof.** See [Abhy90].

On the other hand,

**10.14.9. Example.** A space curve is not necessarily the intersection of two surfaces. Consider, for example, the twisted cubic. Since a plane in general position will intersect this curve in three points, it has degree 3. But Bèzout's theorem then implies that the curve would have to be the intersection of a plane and a cubic surface. This is impossible since it is not a plane curve.

## 10.15 Parameterizing Implicit Curves

Given a rational parametric representation for a space, it is always possible to represent the space in implicit form via equations. We can do this either using the resultant or Gröbner bases techniques. The converse problem is unfortunately not so simple. In fact, there are implicitly defined plane curves that do not admit a representation via rational polynomial functions. What can be said about this problem tends to get quite complicated and so this section will restrict itself to only some of the simpler results.

Theorems 10.13.5 and 10.13.14 are fundamental for this section. They answer the question of when a curve can be parameterized. We must answer:

> When is an extension field K over a field k which has transcendence degree 1 and is generated by two elements isomorphic to the field k(t) of rational functions in one variable?

Actually, the key condition is that the transcendence degree is 1. We could allow the number of generators to be n; however, this would not gain us anything in generality. Once the extension field satisfies the conditions we can get a parameterization. See Theorem 10.13.18.

The basic approach to parameterize a set **X** is to do a central projection from a point **p** not in the set onto a d-dimensional plane. This will give a parameterization of **X** with d coordinates provided that the lines through **p** meet **X** in only one point. Choosing the point **p** so that this will happen is the hard part. The case where the lines meet **X** in a finite number of points is the next best case. It essentially gives us local parameterizations.

Here is another approach to parameterizing plane curves. Consider a conic. If we can, by a linear change in variables of the form

$$X' = \alpha_1 X + \beta_1 Y + \gamma_1 \quad \text{and} \quad Y' = \alpha_2 X + \beta_2 Y + \gamma_2, \tag{10.85}$$

eliminate the quadratic term of one of the variables, say Y, then we could set X = t, and solve the resulting linear equation in Y for Y giving us Y as a function of t also. Specifically, after eliminating the $Y^2$ term in the equation for the conic, we will have an equation of the form

$$(aX + b)Y + \left(cX^2 + dX + e\right) = 0, \tag{10.86}$$

which has solutions

$$\begin{aligned} X &= t \\ Y &= \frac{-\left(ct^2 + dt + e\right)}{at + b}. \end{aligned} \tag{10.87}$$

We can see from this that we always have one "point at infinity" ($t = \infty$), but we may have a second one for $t = -b/a$ if $a \neq 0$. In general, we will get a rational parameterization, but if a = 0 in equation (10.86), then we actually get a polynomial parameterization.

To get a better picture of what is happening, consider a degree n curve for a moment and write

$$f(X, Y) = f_0 + f_1 + \ldots f_n,$$

where the $f_i$ are homogeneous of degree i. The $Y^n$ term that we want to eliminate is part of $f_n$. Note that over the complex numbers

$$f_n(X, Y) = \prod_{i=1}^{d} (a_i Y - b_i X), \tag{10.88}$$

by Proposition 10.5.3, so that $f_n = 0$ corresponds to n lines through origin. The curve will have no $Y^n$ term if and only if some $a_i$ is zero and one of the lines is X = 0, that is, a line parallel to the y-axis meets the curve at infinity. We can see things even better if we switch to homogeneous coordinates. Homogenizing f by replacing X by X/Z and Y by Y/Z gives

$$F(X, Y, Z) = f_0 Z^n + f_1(X, Y) Z^{n-1} + \ldots + f_{n-1}(X, Y) Z^1 + f_n(X, Y). \tag{10.89}$$

Finding the intersection of the (projective) plane curve

$$F(X, Y, Z) = 0$$

with the line at infinity (Z = 0) means solving

$$f_n(X, Y) = 0.$$

In other words, the $[a_i,b_i,0]$ are the intersection points at infinity. The points may be complex (as in case of unit circle). Therefore, we need to look for real solutions. It is easy to check that the change of variables shown in (10.85) will eliminate the $Y^n$ term if and only if $(\beta_1,\beta_2)$ is a multiple of one of the $(a_i,b_i)$. Now, since we are in projective space we do not have to restrict ourselves to the affine linear transformations defined by equations (10.85). On the other hand, projective linear transformations correspond to fractional transformations in the affine world and so it is then even more likely that we may get rational parameterizations rather than polynomial ones.

We got a fairly nice answer for conics. In the case of cubic plane curves, if we can find a double point, then lines through it will intersect curve in a single point and the same approach will work.

**10.15.1. Theorem.** A cubic plane curve admits a rational parameterization if and only if it has a double point (it is a singular curve).

**Proof.** See [Abhy90].

**10.15.2. Theorem.** If a plane curve has more than one point at infinity, then it cannot be parameterized by polynomials.

**Proof.** Assume that

$$c(t) = (p(t), q(t))$$

is a parameterization of the curve with polynomials p(t) and q(t). The affine part of the associated projective curve has parameterization

$$C(t) = [p(t), q(t), 1].$$

Dividing through by $t^d$, where $d = \max(\deg(p), \deg(q))$, and letting t approach $\pm\infty$ shows that the curve has only one point at infinity.

The converse is not true if the degree of the curve is larger than two, but

**10.15.3. Theorem.** A plane curve can be parameterized by polynomials if and only if it can be parameterized by rational functions and has only one place at infinity.

**Proof.** See [Abhy90].

A criterion for when a plane curve has only one place at infinity can be found in [Abhy90].

It is possible to relate the problem of rational parameterizations to its topology. First, let **C** be an irreducible curve of order n and note that Theorem 10.7.8 can be restated as saying that

$$(n-1)(n-2) \geq \sum_i m_i(m_i - 1), \tag{10.90}$$

where $m_i$ is the multiplicity of the ith singularity and we sum over all singularities of **C**. It turns out that the difference between the maximum number of double points of a curve and actual number of double points defines an important invariant.

**Definition.** Let **C** be a plane curve in $\mathbf{C}^2$ of order n defined by f(X,Y) = 0 that has only ordinary singularities. The *genus* g of **C** is defined by

$$g = (n-1)(n-2) - \sum_i m_i(m_i - 1),$$

where $m_i$ is the multiplicity of the ith singularity and we sum over all singularities.

For a general definition of the genus of a curve, even if it has nonregular singularities, see, for example, [Walk50]. Note that inequality (10.90) implies that the genus is a nonnegative integer.

**10.15.4. Theorem.** The genus of an irreducible curve is a birational invariant.

**Proof.** See [Walk50].

Theorem 10.15.4 has lots of consequences. For example, any nonsingular cubic curve in $\mathbf{P}^2$ is not rational because it has genus 1 and the genus of $\mathbf{P}^1$ is 0. One can also conclude the following (see [Walk50] for details):

(1) There are an infinite number of birationally nonequivalent curves. One way to show this is to consider the curves $\mathbf{C}_m$ defined by

$$Y^2 - F(X) = 0, \tag{10.91}$$

where F(X) is any polynomial of odd degree n = 2m + 1 with no multiple roots. The curve $\mathbf{C}_m$ is nonsingular. The nonsingular projective version of $\mathbf{C}_m$ is called a *hyperelliptic curve*. One can show that the hyperelliptic curve has genus m. See [Shaf94].

(2) No nonsingular irreducible plane curves are equivalent if they have different orders, the only exception being the case where one is a line and the other a conic.

In general, curves with the same genus need not be birationally equivalent. There is one case where they are, however.

**10.15.5. Theorem.** (Noether's Theorem) A plane curve admits a rational parameterization (and is birationally equivalent to a line) if and only if it has genus 0.

**Proof.** See [Walk50], [Hoff89], [Abhy90], [Harr92], or [Shaf94].

It follows that any nonsingular cubic curve in $\mathbf{P}^2$ is not rational because it has genus 1.

The genus that we just defined is an algebraic concept, but if we are talking about curves over the complex numbers, then Theorem 10.2.5 implies that the curve is either a surface or a singular surface where a finite number of points have been identified. In this case, the algebraic genus and the topological one which was defined in Chapter 6 are the same. (In the singular surface case consider the topological genus to be defined by equations (6.3).) The proof of Theorem 10.15.5 does not really provide a good algorithm for finding the rational parameterizations. There are simple algorithms in the case of curves of order 2 or 3 or so-called *monoids* (a curve of degree n with a point of multiplicity n – 1). See the discussion earlier in the section or the lengthy discussion in [Hoff89].

The genus g is not the only birational invariant of a curve. There are continuous invariants called "moduli." Only when two curves have genus 0 are they isomorphic. There are curves with the same genus $g > 0$ that are not isomorphic. See [Shaf94].

We finish this section with one final observation. If an implicitly defined set consists of a finite number of points (so that a parameterization amounts to simply listing the zeros), then there is an algorithm that will find those zeros. The algorithm involves using Gröbner bases. See [Mish93].

## 10.16 The Dimension of a Variety

Every topological space has a notion of dimension associated to it. We defined this concept in the case of cell complexes and manifolds. Since an algebraic variety lives in projective space, it can be thought of as a topological space and so has a dimension. The interesting question is whether one can determine its dimension from its algebraic structure and if yes, then how one would compute it.

Here are some simple two-dimensional examples of varieties V(f) in $\mathbf{R}^3$.

**10.16.1. Example.** $f(X,Y,Z) = aX + bY + cZ + d$

**Description.** V(f) is an arbitrary plane that intersects the x-, y-, and z-axis at –d/a, –d/b, and –d/c, respectively (if a, b, and c are nonzero).

**10.16.2. Example.** $f(X,Y,Z) = X^2 + Y^2 + Z^2 - 1$

**Description.** V(f) is the sphere of radius 1 about the origin.

**10.16.3. Example.** $$f(X,Y,Z) = \frac{X^2}{a^2} + \frac{Y^2}{b^2} - \frac{Z^2}{c^2} - 1$$

**Description.** V(f) is a connected hyperboloid with horizontal slices that are ellipses. See Figure 10.21(a).

**10.16.4. Example.** $$f(X,Y,Z) = \frac{X^2}{a^2} - \frac{Y^2}{b^2} - \frac{Z^2}{c^2} - 1$$

**Description.** V(f) is a disconnected hyperboloid with vertical slices that are ellipses. See Figure 10.21(b).

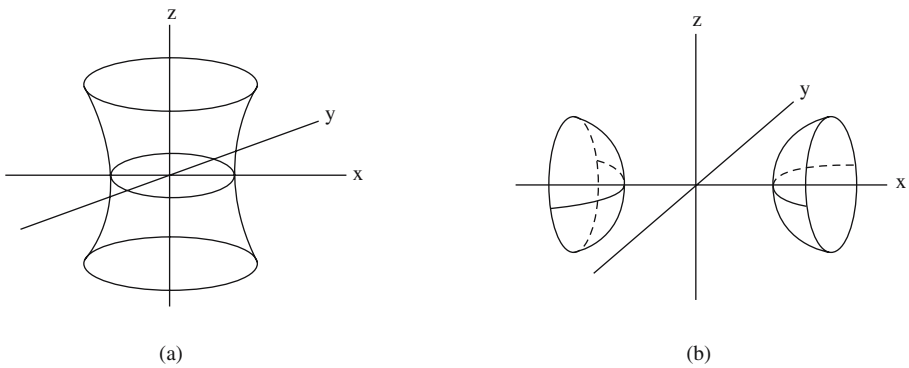



**Figure 10.21.** Some varieties in $\mathbf{R}^3$.

The four examples above are very nice and manifold-like, although, in general, varieties can have bad singularities, even in $\mathbf{R}^3$. Concentrating on the nice cases for a moment, one might be tempted to define dimension in terms of parameterizations like we did for manifolds. In fact, there is a version of the implicit function theorem for complex analytic maps that gives us a criterion for when a point is not a singular point and has a neighborhood that can be parameterized. The precise version of this theorem is too technical to state here, but it has the same flavor as Theorem 4.4.7. See [Kend77] for details. We sketch the basic idea.

Let $\mathbf{V} = V(f_1, f_2, \ldots, f_m)$, $f_i \in \mathbf{C}[X_1, X_2, \ldots, X_n]$, be a variety in $\mathbf{C}^n$. Define

$$F: \mathbf{C}^n \to \mathbf{C}^m$$

by $F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ and consider the Jacobian matrix $F'$ for this map, that is,

$$F' = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}. \tag{10.92}$$

One can show that if $F'$ has rank r in the neighborhood of a point $\mathbf{p}$ in $\mathbf{C}^n$, then a neighborhood of $\mathbf{p}$ **in** $\mathbf{V}$ is an $(n - r)$-dimensional **complex** manifold. Because we are dealing with a local property here, one that only involves points in an arbitrarily small neighborhood of a point, one can prove a similar result for projective varieties since $\mathbf{P}^n(\mathbf{C})$ looks like affine space locally. These results can be used to define the dimension of a variety.

**Definition.** If $\mathbf{p}$ is a point of $\mathbf{V}$ that has a neighborhood in $\mathbf{C}^n$ on which the Jacobian matrix $F'$ in formula (10.92) has constant rank r, then $\mathbf{p}$ is called a *smooth point*

of **V**. The (**complex**) *dimension* of **V** *at* such *a smooth point* is then defined to be n – r and is denoted by $\dim_{\mathbf{p}}(\mathbf{V})$. Similar definitions are made for varieties in $\mathbf{P}^n(\mathbf{C})$. A variety is called a *smooth variety* if every one of its points is a smooth point.

Note that the dimension in the definition is a **complex** dimension, so that as a real manifold, the variety would have **real** dimension $2\dim_{\mathbf{p}}(\mathbf{V})$ at **p**.

**10.16.5. Theorem.** The set of smooth points of a variety in either $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ are dense in the variety.

**Proof.** See [Kend77].

Theorem 10.16.5 implies that the next definition is well defined.

**Definition.** The *dimension* of a variety **V** *at a point* **p** in **V**, denoted by $\dim_{\mathbf{p}}(\mathbf{V})$, is defined by

$$\dim_{\mathbf{p}}(\mathbf{V}) = \lim_{\mathbf{U}_i} \max \{\dim_{\mathbf{q}}(\mathbf{V}) \mid \mathbf{q} \text{ is a smooth point in } \mathbf{U}_i\},$$

where $\mathbf{U}_i$ is any sequence of neighborhoods of **p** that converge to **p**. In other words, $\dim_{\mathbf{p}}(\mathbf{V})$ is the maximum of the dimensions of **V** at smooth points in an arbitrarily small neighborhood of **p**. The *dimension* of **V**, denoted by dim **V**, is defined by

$$\begin{aligned} \dim \mathbf{V} &= -1, && \text{if } \mathbf{V} \text{ is empty,} \\ &= \max \{\dim_{\mathbf{p}}(\mathbf{V}) \mid \mathbf{p} \in \mathbf{V}\}, && \text{otherwise.} \end{aligned}$$

The *codimension* of **V**, denoted by codim **V**, is defined by

$$\operatorname{codim} \mathbf{V} = n - \dim \mathbf{V}.$$

**10.16.6. Theorem.** In an irreducible variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ all points have the same dimension.

**Proof.** See [Kend77].

**Definition.** A variety is said to have *pure dimension* d if it has the same dimension d at each of its points.

**10.16.7. Corollary.** Every irreducible variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ has pure dimension.

**10.16.8. Theorem.** A variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ is a hypersurface if and only if it has pure dimension n – 1.

**Proof.** See [Kend77].

In the next section we shall carry this definition of dimension further to study singularities, intersection multiplicities, and other concepts that we dealt with in the case

of curves. However, we pause at this stage and point out a defect in our definition, at least as far as an algebraic geometer is concerned. Our definition was more topological than algebraic. "Pure" algebraic geometry should not have to rely on a notion of continuity or differentiability of functions.

The next theorem gives several other characterizations of dimension that are purely algebraic in nature. As a result we see that the dimension of a variety does not depend on any particular coordinatization.

**10.16.9. Theorem.** Let $\mathbf{V}$ be a nonempty irreducible algebraic variety in $k^n$ or $\mathbf{P}^n(k)$. If $k = \mathbf{C}$, then there is an integer d with the property that the following statements about $\mathbf{V}$ and d are equivalent:

(1) $\dim \mathbf{V} = d$.
(2) There is a finite map $f: \mathbf{V} \to k^d$.
(3) $\mathrm{tr}_k(k(\mathbf{V})) = d$.
(4) If $\mathbf{V} = \mathbf{V}_0 \supset \mathbf{V}_1 \supset \ldots \supset \mathbf{V}_s \neq \phi$ is any maximal chain of nonempty distinct irreducible subvarieties $\mathbf{V}_i$, then $s = d$.
(5) If $0 = I_0 \subset I_1 \subset \ldots \subset I_s$ is any maximal chain of distinct prime ideals $I_i$ in $k[\mathbf{V}]$, then $s = d$.

The rest of this section is devoted to motivating this theorem and sketching a proof. See [Kend77], [Harr92], or [Shaf94] for more details. It should be noted that as far as proofs are concerned, the only thing important about the complex numbers is that they are algebraically closed. Only the definition of dim $\mathbf{V}$ was inherently dependent on the complex numbers.

(1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3): Consider the affine case. By Theorem 10.13.38(2) there is an m, $m \leq n$ and a finite map

$$f: \mathbf{V} \to k^m.$$

The pullback map

$$f^*: k(k^m) \to k(\mathbf{V})$$

is one-to-one and $k(\mathbf{V})$ can be thought of as a finite extension of $k(k^m)$, which is just $k(X_1, X_2, \ldots, X_m)$. This shows that the transcendence degree of $k(\mathbf{V})$ is m. Since f is a regular map, it follows that $\mathbf{V}$ is locally homeomorphic to $k^m$, so that $m = d$.

Using (3) in Theorem 10.16.9, one can prove the following theorem, which is needed to prove that (4) and (5) are equivalent to (3).

**10.16.10. Theorem.** Let $\mathbf{W}$ be a subvariety of a variety $\mathbf{V}$ in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$.

(1) $\dim \mathbf{W} \leq \dim \mathbf{V}$.
(2) If $\mathbf{V}$ is irreducible and $\dim \mathbf{W} = \dim \mathbf{V}$, then $\mathbf{W} = \mathbf{V}$.

**Proof.** See [Kend77] or [Shaf94].

Another result we shall need is

**10.16.11. Theorem.** Let $\mathbf{V}$ be an irreducible projective variety in $\mathbf{P}^n(k)$ for an algebraically closed field k and assume that $f \in k[X_1, X_2, \ldots, X_{n+1}]$ is a homogeneous polynomial that does not vanish on $\mathbf{V}$. If $\mathbf{V}_f$ is the subvariety of $\mathbf{V}$ defined by the equation $f = 0$, then $\dim \mathbf{V}_f = \dim \mathbf{V} - 1$.

**Proof.** We follow the argument in [Shaf94]. By hypothesis, $\mathbf{V}_f \neq \mathbf{V}$, so that $\dim \mathbf{V}_f < \dim \mathbf{V}$ by Theorem 10.16.10. Set $\mathbf{V}_1 = \mathbf{V}_f$ and $f_0 = f$. Now pick a point in each irreducible component of $\mathbf{V}_1$ and let $f_1$ be a homogeneous polynomial that does not vanish on any of these points. Let $\mathbf{V}_2$ be the subvariety of one of the irreducible components of $\mathbf{V}_1$ defined by the condition that $f_1 = 0$. Continue on in this way to get a sequence of irreducible varieties $\mathbf{V}_i$ and homogeneous polynomials $f_i$ so that

$$\mathbf{V} \supset \mathbf{V}_1 \supset \mathbf{V}_2 \supset \ldots$$

and $\dim \mathbf{V}_{i+1} < \dim \mathbf{V}_i$. Since the dimension of the varieties is decreasing there is a d, $d \leq \dim \mathbf{V}$, so that $\mathbf{V}_d \neq \phi$ and $\mathbf{V}_{d+1} = \phi$. By definition the polynomials $f_0, f_1, \ldots, f_d$ have no common zeros and so we can define a map

$$\varphi: \mathbf{V} \to \mathbf{P}^d(k)$$

by

$$\varphi([\mathbf{x}]) = [f_0(\mathbf{x}), f_1(\mathbf{x}), \ldots, f_d(\mathbf{x})].$$

By Theorem 10.13.37, $\varphi: \mathbf{V} \to \varphi(\mathbf{V})$ is a finite map and by Theorem 10.13.33 $\varphi(\mathbf{V})$ is a subvariety of $\mathbf{P}^d(k)$. It follows that $\dim \mathbf{V} = \dim \varphi(\mathbf{V}) = d$. Note also that $\varphi(\mathbf{V}) = \mathbf{P}^d(k)$ by Theorem 10.16.10. Clearly we must have $\dim \mathbf{V}_i = \dim \mathbf{V}_{i+1} + 1$. In particular, $\dim \mathbf{V}_f = \dim \mathbf{V} - 1$ and the theorem is proved.

Returning to the proof of Theorem 10.16.9, we first observe

(4) $\Leftrightarrow$ (5): This is easy because of the connection between ideals and varieties.

Therefore, to finish the proof of Theorem 10.16.9 we have the choice of proving that (3) is equivalent to (4) or (5).

(3) $\Leftrightarrow$ (4): Theorem 10.16.10 clearly implies that there cannot be a strictly decreasing sequence of irreducible varieties of longer length than d. Repeated application of Theorem 10.16.11 shows that there is at least one of that length in the projective case. In the affine case this follows from an affine version of Theorem 10.16.11. To finish the proof we must show that every maximal such sequence has length d. For that it suffices to show that if $\mathbf{W}_1$ and $\mathbf{W}_2$ are irreducible varieties with $\mathbf{W}_2$ properly contained in $\mathbf{W}_1$ and if

$$\dim \mathbf{W}_2 < \dim \mathbf{W}_1 - 1$$

then there is an irreducible subvariety $\mathbf{W}'$ of $\mathbf{W}_1$ with the property that

$$\mathbf{W}_1 \supset \mathbf{W}' \supset \mathbf{W}_2 \quad \text{and} \quad \dim \mathbf{W}' = \dim \mathbf{W}_1 - 1.$$

To prove this we switch to the ideal version of this fact. We look at the affine case and leave the projective case as an exercise. Specifically, it suffices to show that if I and J are prime ideals in k[**V**] of transcendence degree e and f over k, respectively, with $I \supset J$, then we can find a chain of distinct prime ideals

$$I \supset J_1 \supset \ldots \supset J_{e-f} = J$$

of length e – f. We sketch the proof in [Kend77].

We may assume that

$$k[\mathbf{V}] = \frac{k[X_1, X_2, \ldots, X_n]}{I(V)} = k[x_1, x_2, \ldots, x_n],$$
$$I = k[y_1, y_2, \ldots, y_n],$$
$$J = k[z_1, z_2, \ldots, z_n],$$

where $x_i$ is the projection of $X_i$ in k[**V**], and that $y_1, y_2, \ldots, y_e$ and $z_1 = y_1, z_2 = y_2, \ldots, z_f = y_f$ are a transcendence basis of I and J over k, respectively. Furthermore, we may assume that the natural projection homomorphism

$$\pi : I \to \frac{I}{I(V(J))} = J$$

sends $y_i$ to $z_i$ under the appropriate identifications.

Because the $y_i$ are a transcendental basis for $k[y_1, y_2, \ldots, y_e]$ over k, there is a unique ring homomorphism (over k)

$$\sigma : k[y_1, y_2, \ldots, y_e] \to I.$$

defined by

$$\sigma(y_i) = z_i, \quad 1 \le i \le f+1$$
$$= y_i, \quad f+1 < i \le e.$$

**Claim.** The homomorphism σ extends to a unique ring homomorphism

$$\sigma_1 : I = k[y_1, y_2, \ldots, y_n] \to I.$$

To prove the claim, note that $y_s$ is algebraic over $k[y_1, y_2, \ldots, y_e]$ for $e < s \le n$. Let

$$p_s(y_1, y_2, \ldots, y_e, X) \in k[y_1, y_2, \ldots, y_e][X]$$

be its minimal polynomial. Applying the projection map π, we see that $p_s(z_1, z_2, \ldots, z_e, X)$ has positive degree, which in turn implies that $p_s(z_1, z_2, \ldots, z_{f+1}, y_{f+2}, \ldots, y_e, X)$ has positive degree. The claim now follows from Theorem B.8.15.

Let $J_1 = \sigma_1(I)$. The ideal $J_1$ is prime. Since $z_t$ is algebraic over $k[z_1, z_2, \ldots, z_f]$ for $t > f$, it follows that $J_1$ has transcendence degree e – 1 over k. Repeating the same steps for $J_1$, produces a prime ideal $J_2$, $J_1 \supset J_2$, of transcendence degree e – 2 over k, and so on until we finally get $J_{e-f} = J$.

Theorem 10.16.9 is proved.

Let us expand on the definition of dimension of a projective variety $\mathbf{V} \subseteq \mathbf{P}^n(\mathbf{C})$ in terms of the existence of a finite map

$$f: \mathbf{V} \to k^d$$

(Theorem 10.16.9(2)). Such a map can be obtained by a succession of projections from a point that leads to a projection of $\mathbf{V}$ with center a linear subspace of $\mathbf{P}^n(\mathbf{C})$. It is easy to see that the integer d is characterized by the property that one can find an (n – d – 1)-dimensional plane $\mathbf{X}$ in $\mathbf{P}^n(\mathbf{C})$ (the center of the projection) that does not intersect $\mathbf{V}$, but every (n – d)-dimensional plane in $\mathbf{P}^n(\mathbf{C})$ intersects $\mathbf{V}$.

## 10.17 The Grassmann Varieties

Grassmann manifolds were defined in Section 8.14. We take another look at these special manifolds, but from the point of view of algebraic geometry this time. The Grassmann manifolds $G_n(\mathbf{R}^{n+k})$ can actually be thought of as subvarieties of projective space and are an important example of higher-dimensional varieties. An overview of Grassmann varieties in the context of algebraic geometry can be found in [KleL72]. See also [Harr92] and [Shaf94].

Note that if $\mathbf{V}$ is an n-dimensional linear subspace of $\mathbf{R}^{n+k}$ and if $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is a basis for $\mathbf{V}$, then the wedge product $\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_n$ defines an element in the N-dimensional vector space $\Lambda^n(\mathbf{R}^{n+k})$, where

$$N = \binom{n+k}{n}.$$

We shall identify $\Lambda^n(\mathbf{R}^{n+k})$ with $\mathbf{R}^N$ using the canonical basis

$$\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \ldots \wedge \mathbf{e}_{i_n}, \quad 1 \le i_1 < i_2 < \ldots < i_n \le n+k.$$

Changing the basis B for $\mathbf{V}$ will change the wedge product by a scalar multiple. Therefore, the map

$$\mu: G_n(\mathbf{R}^{n+k}) \to \mathbf{P}^{N-1}$$

defined by

$$\mu(\mathbf{V}) = [\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_n],$$

where $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is a basis for $\mathbf{V}$, is well defined. Now

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \ldots \wedge \mathbf{v}_n = \sum_{1 \le i_1 < i_2 < \ldots i_n \le n+k} a_{i_1 i_2 \ldots i_n} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \ldots \wedge \mathbf{e}_{i_n}.$$

**Definition.** The homogeneous coordinates $a_{i_1 i_2 \ldots i_n}$ are called the *Plücker coordinates* of $\mathbf{V}$.

One can show that the Plücker coordinates are just the $n \times n$ minors of the $n \times (n + k)$ matrix whose rows are the vectors $\mathbf{v}_i$.

**10.17.1. Lemma.** The map $\mu$ is an imbedding.

**Proof.** The lemma follows from the fact that if $\mu(\mathbf{V}) = [\omega]$, then

$$\mathbf{V} = \left\{\mathbf{v} \in \mathbf{R}^{n+k} \mid \mathbf{v} \wedge \omega = 0\right\}.$$

**10.17.2. Theorem.** The set $\mu(G_n(\mathbf{R}^{n+k}))$ is a smooth submanifold of $\mathbf{P}^{N-1}$ and is the set of zeros of a finite set of quadratic homogeneous polynomials called the *Plücker relations*.

**Proof.** See the algebraic geometry references listed above.

Finally, one can generalize the definition of the Grassmann varieties both by allowing a more general field and by passing from an affine to a projective version. In particular, $G_n(\mathbf{C}^{n+k})$ denotes the space of complex n-dimensional linear subspaces of $\mathbf{C}^{n+k}$ and $G_n(\mathbf{P}^{n+k})$ and $G_n(\mathbf{P}^{n+k}(\mathbf{C}))$ denote the spaces of n-dimensional linear subspaces of the projective spaces $\mathbf{P}^{n+k}$ and $\mathbf{P}^{n+k}(\mathbf{C})$, respectively. In all these cases one gets nice varieties that are also manifolds.

## 10.18 N-Dimensional Varieties

This section gives a very brief overview of algebraic geometry in higher dimensions. We continue the approach to higher-dimensional varieties taken in Section 10.16. The results will divide up into two types, those that deal with properties of varieties that are **local** in nature and those that are **global**. For example, some theorems deal with what neighborhoods of points look like (see Theorem 10.18.9). This is a local property. Others deal with whether or not the space is connected, whether it is orientable (in the case where we are dealing with a manifold), or what its homology groups or derived invariants, such as the Euler characteristic, are. These are global properties. In addition to considering intrinsic properties of varieties by themselves, we also want to generalize how intersections behave. We shall see that what we learned about plane curves will generalize if we use the concept of codimension.

Let $\mathbf{V} = V(f_1, f_2, \ldots, f_m)$, $f_i \in \mathbf{C}[X_1, X_2, \ldots, X_n]$, be a variety in $\mathbf{C}^n$. Define

$$F: \mathbf{C}^n \to \mathbf{C}^m$$

by

$$F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots f_m(\mathbf{x})).$$

Here is a local criterion for smoothness.

**10.18.1. Theorem.** If $\mathbf{V}$ is irreducible, then it is smooth at a point $\mathbf{p}$ if and only if

$$\operatorname{codim} \mathbf{V} = \text{rank of Jacobian matrix } F' \text{ at } \mathbf{p}. \tag{10.93}$$

A corresponding result holds for projective varieties in $\mathbf{P}^n(\mathbf{C})$.

**Proof.** See [Kend77].

Theorem 10.18.1 leads to a definition of nonsingularity which generalizes the definition we gave in the case of plane curves.

**Definition.** Let $\mathbf{V}$ be an irreducible variety in $\mathbf{C}^n$ and let $\mathbf{p} \in \mathbf{V}$. If equation (10.93) holds at $\mathbf{p}$, then $\mathbf{p}$ is called a *nonsingular* point of $\mathbf{V}$ and $\mathbf{V}$ is said to be *nonsingular at* $\boldsymbol{p}$. Otherwise, $\mathbf{p}$ is called a *singular* point for $\mathbf{V}$ and $\mathbf{V}$ is said to be *singular at* $\boldsymbol{p}$. If every point of $\mathbf{V}$ is a nonsingular point, then $\mathbf{V}$ is said to be a *nonsingular variety*. A similar definition is made in the case of an irreducible projective variety in $\mathbf{P}^n(\mathbf{C})$.

**10.18.2. Theorem.** The set of singular points of an irreducible variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ form a proper subvariety.

**Proof.** See [Kend77].

Theorems 10.18.1 and 10.18.2 generalize to arbitrary varieties. We now turn to some topological issues.

**10.18.3. Theorem.**

(1) Every algebraic curve in $\mathbf{P}^n(\mathbf{C})$ is connected.
(2) Every irreducible variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ is connected.

**Proof.** See [Kend77]. The reason that irreducibility is not needed in (1) is that the irreducible components of a curve all intersect in this case and the union of connected sets that intersect is connected.

**10.18.4. Theorem.**

(1) Every nonsingular one-dimensional variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ is an orientable surface.
(2) Every irreducible d-dimensional nonsingular variety in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ is orientable as a real 2d-dimensional manifold.

**Proof.** See [Kend77]. Compare (1) with Theorem 10.2.5.

The final topic of this section deals with intersections of varieties and Bèzout's Theorem. There will be a long list of theorems which deal with the intersection of suitable linear subspaces with varieties. The reader should reread the comments at the beginning of Section 10.6 for why this is a reasonable geometric approach to the definitions and theorems we shall state. Intersections of varieties with linear subspaces also played a role in Section 10.7 (in that case we used lines) and in Section 10.16 (see the comments at the end of the section regarding the definition of dimension in terms of finite maps). First, though, it is convenient to state a bound on the dimension of an intersection in terms of its codimension.

**10.18.5. Theorem.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be two irreducible varieties in $\mathbf{C}^n$ that have a **nonempty** intersection. Then

$$\text{codim}\,(\mathbf{V}_1 \cap \mathbf{V}_2) \le \text{codim}\,\mathbf{V}_1 + \text{codim}\,\mathbf{V}_2.$$

**Proof.** See [Kend77].

**10.18.6. Corollary.** If $\mathbf{V}_1$ and $\mathbf{V}_2$ are irreducible varieties in $\mathbf{P}^n(\mathbf{C})$, then

$$\text{codim}\,(\mathbf{V}_1 \cap \mathbf{V}_2) \le \text{codim}\,\mathbf{V}_1 + \text{codim}\,\mathbf{V}_2.$$

**Proof.** This corollary is an immediate consequence of Theorem 10.18.5 because **any** two varieties in $\mathbf{P}^n(\mathbf{C})$ intersect.

**Definition.** We say that two irreducible varieties $\mathbf{V}_1$ and $\mathbf{V}_2$ in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ *intersect properly* if

$$\text{codim}\,(\mathbf{V}_1 \cap \mathbf{V}_2) \le \text{codim}\,\mathbf{V}_1 + \text{codim}\,\mathbf{V}_2.$$

Two arbitrary varieties $\mathbf{V}_1$ and $\mathbf{V}_2$ are said to *intersect properly* if each irreducible component of $\mathbf{V}_1$ intersects properly with each irreducible component of $\mathbf{V}_2$.

The idea of intersecting properly tries to capture the idea that two varieties overlap as little as possible. Too much overlap corresponds to degenerate cases about which not much can be said. The condition on codimension in the definition of intersecting properly is equivalent to saying that the dimension of the intersection is as small as possible, namely, $\dim \mathbf{V}_1 + \dim \mathbf{V}_2 - n$. The ideal case is where varieties intersect transversally but the weaker condition of intersecting properly is adequate. To see that to intersect properly is not the same as intersecting transversally consider the varieties $V(X^2 + Y^2 - 1)$ and $V(Z - 1)$, which intersect properly but not transversally. If varieties intersect transversally, then they will also intersect properly, so that the former condition is stronger than the latter.

In order to state the generalized version of Bèzout's theorem we need to define the degree of an arbitrary variety. There are a number of ways to do this, but first of all, just like in the case of dimension, one has to agree on what the degree should be in simple cases. We already agreed earlier in Section 10.5 that a hypersurface defined by an irreducible polynomial should have degree equal to the degree of that polynomial. A natural way to deal with the general case would be to divide it into two steps:

**Step 1:** Define the degree of an arbitrary irreducible variety (which may not be defined by a single irreducible polynomial since the polynomial ring in more than one variable is not a principal ideal domain).
**Step 2:** Define the degree of an arbitrary variety to be the sum of the degrees of its irreducible components.

Step 2 is plausible given the relationship between the degree of a plane curve and the number of intersections it has with a line discussed at the beginning of Section 10.6. Step 1 is clearly the hard part but we shall deal with it in a similar way, in terms of intersections with linear subspaces.

**10.18.7. Theorem.** Let **V** be a variety of pure dimension d in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$. Almost all transforms of a given (n – d)-dimensional affine or projective plane, respectively, intersect **V** in a common fixed number s of distinct points. If **V** is a hypersurface with minimal polynomial f, then s is the degree of f.

**Proof.** See [Kend77].

**Note.** The set of linear transformations forms a manifold and the expression "almost all transforms" in the last and next several theorems means all transforms except possibly those on a proper lower-dimensional submanifold (a set of measure zero). The planes for which the theorems hold are those that intersect the variety in question transversally. The planes for which the theorems do not hold is a set of measure zero. Another way of looking at the expression "almost all transforms" is via Grassmann manifolds and have it mean "for all n-dimensional linear subspaces in an open dense subset of the Grassmann variety with respect to the Zariski topology."

**Definition.** If **V** is any variety of pure dimension d in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$, then the number s in Theorem 10.18.7 is called the *degree* of **V** and is denoted by deg **V**.

Note that the second part of Theorem 10.18.7 implies that the new definition of the degree of a variety **V** agrees with the definition in Section 10.5 when **V** is a hypersurface.

The degree is a global property. There is a local version of the degree that says that near a point the number of points in the intersection with a linear subspace does not change. We shall describe this also. It is the analog of the multiplicity for plane curves.

**10.18.8. Theorem.** Let **p** be a point of a pure r-dimensional variety **V** in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ and let **L** be any (n – r)-dimensional affine or projective plane in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$, respectively, that intersects **V** properly at **p**. Then for almost all transforms **L′** of **L** sufficiently close to **L**, $\mathbf{V} \cap \mathbf{L}'$ consists of a common fixed number d of points in an arbitrarily small neighborhood of **p**.

**Proof.** See [Kend77].

**Definition.** If **p** is a point of a pure r-dimensional variety **V** in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$, then the number d in Theorem 10.18.8 is called the *multiplicity of intersection of **V** and **L** at **p*** and is denoted by i(**V**,**L**; **p**).

**10.18.9. Theorem.** Let **p** be a point of a pure r-dimensional variety **V** in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$. For almost all transforms **L′** of any (n – r)-dimensional affine or projective plane **L** in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$, respectively, with both **L′** and **L** containing **p**, the number i(**V**,**L′**; **p**) is defined and equal to a common fixed number d. If **V** is a hypersurface with minimal polynomial f, then d is the order of f at **p**.

**Proof.** See [Kend77]. Recall that the order of an arbitrary polynomial at a point is the smallest degree of all the monomials appearing in an expansion about **p**.

**Definition.** If **p** is a point of a pure r-dimensional variety **V** in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$, then the number d in Theorem 10.18.9 is called the *multiplicity of **V** at **p*** or the *order of **V** at **p*** and is denoted by $m_{\mathbf{p}}(\mathbf{V})$.

Note that for a hypersurface **V** the multiplicity of **V** at **p** is $\mathrm{ord}_{\mathbf{p}}(f)$, just like for plane curves.

**10.18.10. Theorem.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ of pure dimension r and s, respectively, and let **L** be any (2n – r – s)-dimensional plane in $\mathbf{P}^n(\mathbf{C})$. For almost all transforms $\mathbf{V}_1'$, $\mathbf{V}_2'$, and $\mathbf{L}'$ of $\mathbf{V}_1$, $\mathbf{V}_2$, and **L**, respectively, $(\mathbf{V}_1' \cap \mathbf{V}_2') \cap \mathbf{L}'$ consists of a common fixed number d of points.

**Proof.** See [Kend77].

The value $2n - r - s = (n - r) + (n - s)$ in the theorem comes from the fact that this is the codimension of $\mathbf{V}_1 \cap \mathbf{V}_2$ if they intersect transversally.

**Definition.** If $\mathbf{V}_1$ and $\mathbf{V}_2$ are any two pure dimensional varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ that intersect properly, then the number d in Theorem 10.18.10 is called the *degree of intersection* of $\mathbf{V}_1$ and $\mathbf{V}_2$ and is denoted by $\deg(\mathbf{V}_1 \bullet \mathbf{V}_2)$.

**Note.** The degree of intersection, $\deg(\mathbf{V}_1 \bullet \mathbf{V}_2)$, is in general **not** equal to $\deg(\mathbf{V}_1 \cap \mathbf{V}_2)$, the degree of the intersection of the two varieties, as one will see in Example 10.18.15 below. The two degrees **are** the same if the varieties intersect transversally.

We have all the definitions needed to state Bèzout's theorem; however, it is worthwhile to show how they relate to a generalized concept of multiplicity of intersections. In analogy with the plane curve case, this concept is introduced by considering intersections of linear subspaces with varieties.

**10.18.11. Theorem.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ of pure dimension r and s, respectively, and let **L** be any (2n – r – s)-dimensional linear subspace. If $\mathbf{V}_1$, $\mathbf{V}_2$, and **L** intersect properly at a point **p**, then for almost all transforms $\mathbf{V}_1'$ of $\mathbf{V}_1$ near $\mathbf{V}_1$, $\mathbf{V}_2'$ of $\mathbf{V}_2$ near $\mathbf{V}_2$, and $\mathbf{L}'$ of **L** near **L** respectively, there is a common fixed number d of distinct points of $\mathbf{V}_1' \cap \mathbf{V}_2' \cap \mathbf{L}'$ near **p**.

**Proof.** See [Kend77].

**Definition.** The fixed number d in Theorem 10.18.11 is called the *intersection multiplicity* of $\mathbf{V}_1$, $\mathbf{V}_2$, and **L** *at **p*** and is denoted by $i(\mathbf{V}_1,\mathbf{V}_2,\mathbf{L};\mathbf{p})$.

**10.18.12. Theorem.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ of pure dimension r and s, respectively, which intersect properly. Then for almost all transforms $\mathbf{L}'$ of a linear subspace **L** of dimension 2n – r – s that contain a point **p**, the number $i(\mathbf{V}_1,\mathbf{V}_2,\mathbf{L}';\mathbf{p})$ is defined and has a common fixed value.

**Proof.** See [Kend77].

**Definition.** The fixed number in Theorem 10.18.12 is called the *intersection multiplicity* of $\mathbf{V}_1$ and $\mathbf{V}_2$ *at* $\boldsymbol{p}$ and is denoted by $i(\mathbf{V}_1,\mathbf{V}_2;\mathbf{p})$.

**10.18.13. Theorem.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be any two pure dimensional varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ which intersect properly. If $\mathbf{C}$ is an irreducible component of $\mathbf{V}_1 \cap \mathbf{V}_2$, then at almost every point $\mathbf{p}$ in $\mathbf{C}$, $i(\mathbf{V}_1,\mathbf{V}_2;\mathbf{p})$ has a common fixed value.

**Proof.** See [Kend77].

**Definition.** The fixed number in Theorem 10.18.13 is called the *intersection multiplicity* of $\mathbf{V}_1$ and $\mathbf{V}_2$ *along* $\boldsymbol{C}$ and is denoted by $i(\mathbf{V}_1,\mathbf{V}_2;\mathbf{C})$.

Finally,

**Definition.** Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be any two pure dimensional varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ that intersect properly. The formal sum

$$\sum_{i=1}^{m} i(\mathbf{V}_i, \mathbf{V}_2; \mathbf{C}_i)\mathbf{C}_i$$

over all the distinct irreducible components $\mathbf{C}_i$ of $\mathbf{V}_1 \cap \mathbf{V}_2$ is called the *intersection product* of $\mathbf{V}_1$ and $\mathbf{V}_2$ and is denoted by $\mathbf{V}_1 \bullet \mathbf{V}_2$.

Note the purely formal nature of the intersection product $\mathbf{V}_1 \bullet \mathbf{V}_2$ and its similarity between the formal sum that it is and the formal sums that are used to define homology groups. We can now tie together the concepts of intersection degree and intersection multiplicities, namely,

**10.18.14. Theorem.** If $\mathbf{V}_1$ and $\mathbf{V}_2$ are any two pure dimensional varieties in $\mathbf{C}^n$ or $\mathbf{P}^n(\mathbf{C})$ that intersect properly, then

$$\deg(\mathbf{V}_1 \bullet \mathbf{V}_2) = \sum_{i=1}^{m} i(\mathbf{V}_1, \mathbf{V}_2, \mathbf{C}_i)\deg \mathbf{C}_i.$$

where the sum on the right-hand side of this equation is taken over all the distinct irreducible components $\mathbf{C}_i$ of $\mathbf{V}_1 \cap \mathbf{V}_2$.

**Proof.** See [Kend77].

Note that the right hand side of the equation in Theorem 10.18.14 would be the natural definition for the degree of the intersection product $\mathbf{V}_1 \bullet \mathbf{V}_2$ although we do not bother to make such a definition.

**10.18.15. Example.** Let $\mathbf{T}$ be the torus in $\mathbf{R}^3$ obtained by rotating the circle in the x-z plane with center (r,0,0) and radius s, where $r > s > 0$, about the z-axis.

**Analysis.** As a surface of revolution, it is easy to see that $\mathbf{T}$ is the **real** hypersurface V(f), where

$$f(X,Y,Z) = \left(X^2 + Y^2 + Z^2 + r^2 - s^2\right)^2 - 4r^2\left(X^2 + Y^2\right).$$

The torus **T** intersects the plane **P** defined by Z = s in a circle defined by

$$X^2 + Y^2 - r^2 = 0.$$

Note that **P** = V(Z – s). Note also that **T** and **X** intersect properly but not transversally.

Now consider everything as defined over the complex numbers. The corresponding **complex** varieties V(f) and V(Z – s) have degree 4 and 2, respectively. The complex variety

$$\mathbf{S} = V(f) \cap V(Z - s)$$

has degree 2. It is also easy to see that

$$i(V(f), V(Z - s); \mathbf{p}) = 2$$

for all points **p** in **S**, and so

$$i(V(f), V(Z - s); \mathbf{S}) = 2.$$

It follows that

$$V(f) \bullet V(Z - s) = 2\mathbf{S},$$

and

$$\deg(V(f) \bullet V(Z - s)) = 2 \deg(V(f) \cap V(Z - s)) = 4.$$

We can at last state Bèzout's theorem.

**10.18.16. Theorem.** (Bézout's Theorem) If two pure dimensional varieties **V** and **W** in $\mathbf{P}^n(\mathbf{C})$ intersect properly, then

$$\deg(\mathbf{V} \bullet \mathbf{W}) = (\deg\ \mathbf{V})(\deg\ \mathbf{W}).$$

If they intersect transversally, then

$$\deg(\mathbf{V} \cap \mathbf{W}) = (\deg\ \mathbf{V})(\deg\ \mathbf{W}).$$

**Proof.** See [Kend77].

**10.18.17. Corollary.** If two plane curves in $\mathbf{P}^2(\mathbf{C})$ of degree m and n intersect in more than nm points counted with their multiplicity, then they must have at least one irreducible component in common.

There is an alternate approach to the degree of a variety and Bèzout's theorem that connects these ideas to homology theory and topology. Suppose that **V** is a

complex variety of $\mathbf{P}^n(\mathbf{C})$. If $\mathbf{V}$ is a complex manifold of dimension 2k, then $\mathbf{V}$ is compact and orientable and a fundamental homology class of $\mathbf{V}$ determines a homology class in $H_{2k}(\mathbf{P}^n(\mathbf{C}))$, which we shall denote by $[\mathbf{V}]$. In fact, an **arbitrary** projective variety determines such a class $[\mathbf{V}]$ not just one that has no singularities. But the topology of $\mathbf{P}^n(\mathbf{C})$ is well-understood and it is known that $H_{2k}(\mathbf{P}^n(\mathbf{C}))$ is isomorphic to $\mathbf{Z}$. Therefore, if $\sigma$ be a generator, then there is a d so that $[\mathbf{V}] = d\sigma$. One can show that $\deg \mathbf{V} = |d|$. Finally, there is a close connection between the degree of intersection of two varieties and their topological intersection numbers, with the final result that Bèzout's theorem can be proved via algebraic topology. We refer reader to [BriK81] and [Harr92].

Finite maps are almost imbeddings. This leads to the question as to whether we can find finite maps which are imbeddings and how large the n has to be. The next theorem is the analog of the Whitney imbedding theorem for differentiable manifolds in the algebraic setting (see the comments following Theorem 8.8.7).

**10.18.18. Theorem.** A nonsingular projective n-dimensional variety is isomorphic to a subvariety of $\mathbf{P}^{2n+1}$.

**Proof.** See [Shaf94].

Every nonsingular "quasi-projective" curve is isomorphic to a curve in $\mathbf{P}^3$. Recall from the discussion after Theorem 10.14.7 that not every curve is isomorphic to a nonsingular one in $\mathbf{P}^2$.

The last result of this section returns to the subject of resolution of singularities and blowups. We looked at an aspect of this for curves in Section 10.12. Blowups are basically a special type of birational map. They are important in the study of rational maps. Lack of space prevents us from going into any details here. The reader is referred to [Harr92] and [Shaf94]. [Harr92] also discusses the following theorem of H. Hironaka (his proof assumes a field of characteristic zero with the general case not yet known):

**10.18.19. Theorem.** For any variety $\mathbf{V}$ we can find a smooth variety $\mathbf{W}$ and a regular birational map $\varphi: \mathbf{W} \to \mathbf{V}$. The map $\varphi$ is called a *resolution of the singularities* of $\mathbf{V}$.

This brief overview of algebraic geometry in higher dimensions consisted mainly of a collection of definitions and unproved theorems but hopefully it gave the reader at least a slight idea of the subject. If the reader is left with the feeling excessive abstractness, of theorems that were true because the definitions were formulated in such a way as to make them true, then this is quite understandable. This is not the only place in mathematics where one encounters such a phenomenon. It definitely does not mean that it is all abstract nonsense though, because one gets concrete results at the end. As we have said before, it is coming up with the right definitions that pick out the essential aspects of a problem that often leads to a breakthrough in the subject and makes everything seem simple to prove afterwards.

In closing we should mention the following. We have studied varieties as subsets of $k^n$ or $\mathbf{P}^n(k)$, $k = \mathbf{R}$ or $\mathbf{C}$, but one can also study them intrinsically the way one studies abstract manifolds. They can be given intrinsic differentiable and analytic structures and one can do calculus on them.

## 10.19 EXERCISES

### Section 10.1

10.1.1. Show that $\mathbf{C} - \{\mathbf{0}\}$ cannot be a variety in $\mathbf{C}$.

10.1.2. Show that a proper nonempty variety in $\mathbf{C}$ consists of a finite set of points.

### Section 10.2

10.2.1. Show that the closure in $\mathbf{P}^2(\mathbf{C})$ of the variety in Example 10.2.3 really is a sphere.

### Section 10.3

10.3.1. Prove Proposition 10.3.1.

10.3.2. Find coordinate neighborhoods $(\mathbf{U},\varphi)$ for $\mathbf{P}^2$ so that the lines defined by the equations below becomes their lines at infinity:

(a) $Y = 0$
(b) $X - 3Y = 0$

10.3.3. Find coordinate neighborhoods $(\mathbf{U},\varphi)$ for $\mathbf{P}^3$ so that the planes defined by the equations below becomes their planes at infinity:

(a) $W = 0$
(b) $X + Y + Z + W = 0$

10.3.4. Consider the polynomial

$$g(X,Y) = Y^2X + X - 1$$

and let $\mathbf{W} = V(g) \subseteq \mathbf{R}^2$. Show that the topological closure of $\mathbf{W}$ in $\mathbf{P}^2$ is $\mathbf{W} \cup \{[0,1,0]\}$.

10.3.5. Consider the polynomial

$$f(X,Y) = Y^2 - X^2(X-1)$$

and let $\mathbf{X}$ be the projective variety in $\mathbf{P}^2(\mathbf{C})$ defined by $H(f)$. Show that topologically $\mathbf{X}$ is a pinched sphere. (**Hint:** We can relate this problem to the one in Example 10.2.4, which considered the variety

$$V\big(Y^2 - (X+\varepsilon)X(X-1)\big)$$

for $\varepsilon = 1$. If we let $\varepsilon$ go to zero, then we will get the variety $\mathbf{V} = V(f) \subseteq \mathbf{R}^2$. In Figure 10.4 (a) the circle will shrink to a point.)

10.3.6. Find a coordinate system for $\mathbf{P}^2$ that shows that the projective completion of the curve $XY = 1$ in $\mathbf{R}^2$ is an ellipse.

10.3.7. (a) Show that for any finite set of points in $\mathbf{P}^2$, there is a line in $\mathbf{P}^2$ that does not pass through these points.

(b) Generalize (a) and show that for any finite set of points in $\mathbf{P}^n$, there is a hyperplane in $\mathbf{P}^n$ that does not pass through these points.

### Section 10.4

10.4.1. Consider the polynomial $f(X) = (X - 1)^2$. Show that $R(f,f') = 0$, thereby verifying Corollary 10.4.5 in this case.

10.4.2. Prove Corollary 10.4.10.

### Section 10.5

10.5.1. Show that the only varieties in $\mathbf{C}^2$ are $\mathbf{C}^2$, a finite (possibly empty) set of points, or the union of a plane curve and a finite (possibly empty) set of points. (**Hint:** Suppose that a variety $\mathbf{X}$ is defined by polynomial equations

$$f_1 = f_2 = \ldots = f_k = 0.$$

Let g be the greatest common factor of the $f_i$ and let $f_i = h_i g$. Determine the conditions under which $\mathbf{X}$ is the union of two varieties, one of which is determined by g and the other by the $h_i$. Use Theorem 10.5.6.)

10.5.2. (a) Show that $f(X,Y) = X^2 + Y^2 - r^2$, $r \in \mathbf{R}$, $r \neq 0$, is irreducible in $\mathbf{C}[X,Y]$ and hence also in $\mathbf{R}[X,Y]$. (**Hint:** Show that any factorization of $f(X,Y)$ must consist of two linear factors and then use Exercise 1.5.19(b).)

(b) Show that $f(X,Y) = X^2 + Y^2 + r^2$, $r \in \mathbf{R}$, $r \neq 0$, is irreducible in $\mathbf{C}[X,Y]$. (**Hint:** Consider the transformation $(x,y) \to (\mathbf{i}x,\mathbf{i}y)$ in $\mathbf{C}^2$.)

10.5.3. Look back at Example 10.3.7 where we considered the polynomial

$$g(X,Y) = Y^2X + X - 1,$$

and the variety $\mathbf{W} = V(g) \subseteq \mathbf{R}^2$. Show the following

(a) $H(g)$ is irreducible in $\mathbf{R}[X,Y,Z]$.

(b) The projective completion $H(\mathbf{W})$ in $\mathbf{P}^2$ is just $V(H(g))$.

10.5.4. Just because a set is described by transcendental functions does not automatically mean that it is not an algebraic variety. Consider, for example, the circle $\{ (\cos t, \sin t) \mid t \in \mathbf{R}\}$. On the other hand, show that the graph of the sine function, $\{ (t, \sin t) \mid t \in \mathbf{R}\}$, is not an algebraic variety.

### Section 10.6

10.6.1. Consider the polynomial

$$f(X,Y) = X^2Y - XY - 2X^2 + Y^2 + 2X - 4Y + 4.$$

Compute $\deg_{(1,2)} f$.

10.6.2. Consider the curve $V(Y - X^2)$ in $\mathbf{R}^2$. Using only the definition, find the multiplicity of the curve at $(-1,1)$ and its tangent line.

10.6.3. Draw the plane curves below and analyze them at the point $\mathbf{p}$ like we did in Examples 10.6.6–10.6.10.

(a) $V(X^2 + Y^2 - 2Y)$, $\mathbf{p} = (1,1)$
(b) $V(X^3 - Y^4)$, $\mathbf{p} = (0,0)$
(c) $V((X^2 + Y^2)^2 + 3X^2Y - Y^3)$, $\mathbf{p} = (0,0)$

10.6.4. Show that any curve in $\mathbf{C}^2$ of degree n that has a point $\mathbf{p}$ of multiplicity n consists of n (not necessarily distinct) lines through $\mathbf{p}$.

### Section 10.7

10.7.1. Consider the curves $\mathbf{C}_1 = V(YZ - X^2)$ and $\mathbf{C}_2 = V(Y^2 - XZ)$ in $\mathbf{P}^2(\mathbf{C})$. Compute the intersection multiplicities $m_{\mathbf{p}}(\mathbf{C}_1,\mathbf{C}_2)$ at the intersection points $\mathbf{p}$ and verify the validity of Bézout's theorem

### Section 10.8

10.8.1. Prove that the graph of any polynomial function f(x,y) in $\mathbf{R}^3$ is an irreducible variety.

10.8.2. Show that

$$<X^2, XY> = <X> \cap <X^2, Y> = <X> \cap <X^2, X + Y>$$

are two irredundant intersections of irreducible ideals. Therefore the direct analog of Theorem 10.8.16 for ideals fails. The closest we come is Theorem B.6.8. By this theorem and Lemma B.6.7 only the associated intersection into prime ideals is unique.

10.8.3. Prove that the radical of a primary ideal in a commutative ring is prime.

10.8.4. Prove that a prime ideal in a commutative ring is irreducible.

10.8.5. This exercise describes another approach to projective varieties. Assume that the field k is infinite in this exercise.

**Definition.** We say that a polynomial $f \in k[X_1,X_2, \ldots ,X_{n+1}]$ *vanishes* at a point $\mathbf{p} \in \mathbf{P}^n(k)$ if

$$f(c_1, c_2, \ldots, c_{n+1}) = 0 \text{ for all } (c_1, c_2, \ldots c_{n+1}) \text{ with } \mathbf{p} = [c_1, c_2, \ldots, c_{n+1}].$$

Given a set S of polynomials in $k[X_1,X_2, \ldots ,X_{n+1}]$, define

$$V(S) = \{\mathbf{p} \in \mathbf{P}^n(k) \mid \text{every } f \in S \text{ vanishes at } \mathbf{p}\}.$$

Any such subset of $\mathbf{P}^n(k)$ is called a *projective variety*.

(a) Let $f \in k[X_1,X_2, \ldots ,X_{n+1}]$. Show that if

$$f = f_0 + f_1 + \ldots + f_d,$$

where $f_i$ is the homogeneous components of degree i for f, then f vanishes at a point $\mathbf{p} \in \mathbf{P}^n(k)$ if and only if each $f_i$ vanishes at $\mathbf{p}$. (**Hint:** For fixed $(c_1, c_2, \ldots, c_{n+1})$ consider

$$g(t) = f(tc_1, tc_2, \ldots, tc_{n+1}) = f_0 + tf_1 + \ldots + t^d f_d.)$$

(b) Let S be a set of polynomials in $k[X_1, X_2, \ldots, X_{n+1}]$. Prove that there are a finite number of homogeneous polynomials $f_1, f_2, \ldots, f_s \in k[X_1, X_2, \ldots, X_{n+1}]$, so that

$$V(S) = V(\{f_1, f_2, \ldots, f_s\}).$$

**Definition.** An ideal $I \subseteq k[X_1, X_2, \ldots, X_{n+1}]$ is said to be *homogeneous* if for all $f \in I$ all the homogeneous components of f belong to I.

(c) Prove that an ideal $I \subseteq k[X_1, X_2, \ldots, X_{n+1}]$ is homogeneous if and only if I is generated by a finite set of homogeneous polynomials.

**Definition.** Let $\mathbf{V}$ be a projective variety in $\mathbf{P}^n(k)$. Define

$$I(\mathbf{V}) = \{f \in k[X_1, X_2, \ldots, X_{n+1}] \mid f \text{ vanishes at every point of } \mathbf{V}\}.$$

(d) Prove that $I(\mathbf{V})$ is a homogeneous ideal.

It follows from (a)–(d) that the definitions in this exercise agree with the corresponding definitions in this chapter.

10.8.6. This exercise deals with some properties of the Zariski topology.

(a) Show that if $\mathbf{U}_1$ and $\mathbf{U}_2$ are two nonempty open subsets of a variety $\mathbf{V}$, then $\mathbf{U}_1 \cap \mathbf{U}_2 \neq \phi$. It follows that the Zarisk i topology is not Hausdorff. Also, every open subset of a variety is dense in it.

(b) Show that any infinite subset of a plane curve is dense in the curve.

(c) Let $\mathbf{V}$ be a variety in $k^n$. Using the notation defined by equations (10.23), show that the projective completion of $\mathbf{V}$ in $\mathbf{P}^n(k)$ relative the coordinate system $(\mathbf{U}_{n+1}, \varphi_{n+1})$ is the closure (in the Zariski topology) of $\varphi_{n+1}^{-1}(\mathbf{V})$ in $\mathbf{P}^n(k)$.

10.8.7. (a) Let $f \in k[X_1, X_2, \ldots, X_n]$. Show that if $I = \langle f \rangle$, then $H(I) = \langle H(f) \rangle$.

(b) Consider the variety $\mathbf{V} = V(Y - X^2, Z - X^3)$ in $k^3$. Show that

$$I = I(\mathbf{V}) = \langle Y - X^2, Z - X^3 \rangle \quad \text{and} \quad ZW - XY \in H(I) \subset k[X, Y, Z, W],$$

but

$$ZW - XY \notin \langle H(Y - X^2), H(Z - X^3) \rangle.$$

This example from [Fult69] shows that (a) does not generalize, namely, if $I = \langle f_1, f_2, \ldots, f_s \rangle$, then it is not necessarily true that $H(I) = \langle H(f_1), H(f_2), \ldots, H(f_s) \rangle$.

## Section 10.9

10.9.1. Find an implicit equation for the parameterized curve

$$x = t + t^2$$
$$y = -t + t^2$$

using the resultant like in Examples 10.9.1. Compare your answer with what you would get by simple elimination of t in the equations:

**Section 10.10**

10.10.1. List the following monomials in three variables $X_1$, $X_2$, and $X_3$ in the degrevlex order:

$$X_2^3, X_1X_2X_3, X_1^2X_3, X_2X_3^2, X_1^3, X_2^3.$$

10.10.2. Show that the degrevlex and deglex order are the same in the case of two variables.

10.10.3. Apply Algorithm 10.10.2 to the polynomials

(a) $f = X^2 + 2XY^2 - XY$, $p_1 = 3X + Y - 1$
(b) $f = X^2Y + 1$, $p_1 = X^2 + X$, $p_2 = XY + X$

Use the deglex order and assume that $Y < X$.

10.10.4. If $g(X_1,X_2, \ldots ,X_n) \in k[X_1,X_2, \ldots ,X_n]$ and $\alpha_1, \alpha_2, \ldots , \alpha_n \in k$, show that

$$g(X_1X_2,\ldots,X_n) = \sum_{i=1}^{n} h_i(X_1,X_2,\ldots,X_n)(X_i - \alpha_1) + g(\alpha_1,\alpha_2,\ldots,\alpha_n)$$

for $h_i(X_1,X_2, \ldots ,X_n) \in k[X_1,X_2, \ldots ,X_n]$.

10.10.5. Let

$$f = X^3Y + X^2Y^2 + Y^2 + Y,$$
$$P = \{ p_1 = X^3 + X, p_2 = X^2 + XY, p_3 = XY - Y \}.$$

Find a P-normal form for f with respect to the deglex order assuming that $Y < X$.

10.10.6. Consider the polynomials

$$f = XY^2 - X, \quad p_1 = XY - Y, \quad p_2 = Y^2 - X \in \mathbf{R}[X,Y].$$

Let $P = \{p_1,p_2\}$. Show that

$$f \xrightarrow{P} 0 \quad \text{and} \quad f \xrightarrow{P} X^2 - X$$

with respect to the deglex order assuming that $Y < X$. Since $f = Yp_1 + p_2$ belongs to the ideal <P> in $\mathbf{R}[X,Y]$, this shows that the mere fact that a polynomial belongs to the ideal <P> does not guarantee that every one of its P-normal forms is zero.

10.10.7. Use Theorem 10.10.12 to determine which of the following sets of polynomials P are Gröbner bases for the ideal I, if any, with respect to the deglex order assuming that $Y < X$:

(a) $P = \{ p_1 = XY - Y, p_2 = Y^2 - X \}$
(b) $P = \{ p_1 = X^2 + X, p_2 = XY + Y, p_3 = Y^2 + Y \}$

10.10.8. Use the deglex order on k[X,Y] and Algorithm 10.10.13 to find Gröbner bases for the ideals <P> below assuming that Y < X:

(a) $P = \{ XY + X, X^2 + Y \}$
(b) $P = \{ X^2Y + X, X + Y \}$

10.10.9. Consider the ideal $I = \langle X^2Y - X - Y, XY^2 + Y \rangle$ in k[X,Y]. Use a Gröbner basis to determine which, if any, of the polynomials below belongs to I. If it does, then express the polynomial in terms of that Gröbner basis.

(a) $f = X^3Y + 2X^2Y^2 + XY^3 - X^2 - XY$
(b) $f = X^3Y + X^2Y^2 - XY + X^2Y - XY^2 - X^2$

10.10.10. Solve Exercise 10.9.1 using Gröbner bases.

**Section 10.12**

10.12.1. Let **C** be a plane curve in $\mathbf{P}^2(k)$. Show that a parameterization of **C** defined in one coordinate system will remain a parameterization when transformed to another coordinate system.

10.12.2. Show that

$$\gamma(t) = \left[\frac{3-3t^2}{1+t^2}, \frac{4t}{1+t^2}, 1\right]$$

is a parameterization of the projective curve in $\mathbf{P}^2(\mathbf{C})$ defined by

$$4X^2 + 9Y^2 - 36Z^2 = 0.$$

Find its center.

10.12.3. Let $\gamma(t)$ be the parameterization in Exercise 10.12.2. Let $h(t) = t + t^2$. Show by direct computation that $\gamma_h(t) = \gamma(h(t))$ has the same center as $\gamma(t)$.

10.12.4. Consider the irreducible curves below:

(a) $Y^2 - X^5 = 0$
(b) $X^4 + X^2Y^2 - Y^2 = 0$

(One way to see that this curve is irreducible is to note that it has a parameterization

$$\gamma(t) = \left(\frac{t^2-1}{t^2+1}, \frac{t^4-2t^2-1}{2t(t^2+1)}\right).)$$

What are their singular points? Find a sequence of quadratic transformations that transform them into curves with only ordinary singularities.

**Section 10.13**

10.13.1. Consider the affine conic defined by

$$f(X,Y) = 5X^2 - 6XY + 5Y^2 - 14X + 2Y + 5 = 0.$$

Find a parameterization of the curve using the method described in Example 10.13.1 and the point (3/2,1/2) on the curve.

10.13.2. Prove Theorem 10.13.2(2).

10.13.3. (a) If $f \in k[\mathbf{V}]$ is a polynomial function on an affine variety $\mathbf{V}$, then $f: \mathbf{V} \to k$ is a continuous function with respect to the Zariski topology.
(b) Generalize (a) and prove that any polynomial function between affine varieties is continuous.

10.13.4. Prove or disprove that the following maps define isomorphisms:

(a) $f: V(XY - 1) \to \mathbf{R}$, $f(x,y) = x$
(b) $g: \mathbf{R} \to V(Y^3 - X^4)$, $g(t) = (t^3, t^4)$
(c) $h: \mathbf{R} \to V(Y - X^k)$, $h(t) = (t, t^k)$

10.13.5. Let $\mathbf{X}$ and $\mathbf{Y}$ be varieties in $k^n$. Let $\Delta = \{ (\mathbf{v},\mathbf{v}) \mid \mathbf{v} \in k^n\}$ be the diagonal in $k^{2n} = k^n \times k^n$.

(a) Show that $\mathbf{X} \times \mathbf{Y}$ and $\Delta$ are varieties in $k^{2n}$.
(b) Define

$$\varphi: \mathbf{X} \cap \mathbf{Y} \to k^{2n} \quad \text{by} \quad \varphi(\mathbf{v}) = (\mathbf{v}, \mathbf{v}).$$

Show that $\varphi$ defines an isomorphism between $\mathbf{X} \cap \mathbf{Y}$ and $(\mathbf{X} \times \mathbf{Y}) \cap \Delta$. In other words, one can replace an intersection between varieties with the intersection of another variety and a linear variety.

10.13.6. Show that a rational function $u: \mathbf{V} \to \mathbf{W}$ between varieties $\mathbf{V}$ and $\mathbf{W}$ is dominant if and only if $\mathbf{W}$ is the smallest variety in $\mathbf{W}$ containing $u(\mathbf{V})$.

10.13.7. Let $f(X,Y) = X^3 - X^2 + Y^2$ and $g(X,Y) = X^2 + Y^2 + X$. Show that the map

$$\varphi(X, Y) = \left( \frac{X^2}{X^2 + Y^2}, \frac{XY}{X^2 + Y^2} \right)$$

sends the variety $V(f)$ to the variety $V(g)$. Show also that the two places of f with center 0 are mapped to places of g with distinct centers. See Figure 10.22.

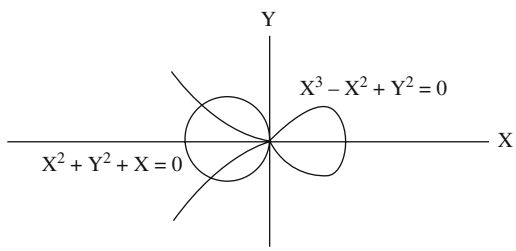



**Figure 10.22.** The curves in Exercise 10.13.7.

10.13.8. Show that the map f in Example 10.13.31 cannot be expressed as a pair of homogeneous polynomials without common zeros.

**Section 10.15**

10.15.1. What is the genus of the cubic $f(X,Y) = X^3 - Y^2$?

APPENDIX A

# Notation

| | |
|---|---|
| $\mathbf{N}$ | = the natural numbers $\{0,1,2,\ldots\}$ |
| $\mathbf{Z}$ | = the ring of integers |
| $\mathbf{Q}$ | = the field of rational numbers |
| $\mathbf{R}$ | = the field of real numbers |
| $\mathbf{R}^*$ | = the extended real numbers, that is, $\mathbf{R} \cup \{\infty\}$ |
| $\mathbf{I}$ | = the unit interval $[0,1]$ |
| $\mathbf{C}$ | = the field of complex numbers |
| $\mathbf{H}$ | = the noncommutative division ring of quaternions |

In the context of an n-tuple $\mathbf{p}$, $p_i$ will always refer to the ith component of $\mathbf{p}$. The same holds for functions. If $f: \mathbf{R}^n \to \mathbf{R}^m$, then $f_i$ is the ith component function of f, that is,

$$f(\mathbf{p}) = (f_1(\mathbf{p}), f_2(\mathbf{p}), \ldots, f_m(\mathbf{p})).$$

| | |
|---|---|
| $\mathbf{N}^n$ | $= \{\mathbf{z} = (z_1,z_2,\ldots,z_n) \mid z_i \in \mathbf{N}\}$ |
| $\mathbf{Z}^n$ | $= \{\mathbf{z} = (z_1,z_2,\ldots,z_n) \mid z_i \in \mathbf{Z}\}$ |
| $\mathbf{R}^n$ | $= \{\mathbf{p} = (p_1,p_2,\ldots,p_n) \mid p_i \in \mathbf{R}\}$ |
| | = n-dimensional Euclidean space |
| $\mathbf{R}^n_+$ | $= \{\mathbf{p} \in \mathbf{R}^n \mid p_n \geq 0\}$ |
| | = the upper halfplane of $\mathbf{R}^n$ |
| $\mathbf{R}^n_-$ | $= \{\mathbf{p} \in \mathbf{R}^n \mid p_n \leq 0\}$ |
| | = the lower halfplane of $\mathbf{R}^n$ |
| $\mathbf{I}^n$ | $= \{\mathbf{p} = (p_1,p_2,\ldots,p_n) \mid 0 \leq p_i \leq 1\}$ |
| | = the unit "cube" in $\mathbf{R}^n$ |
| $\delta_{ij}$ | = Kronecker delta (1, if i = j, and 0, otherwise) |
| $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ | = standard (orthonormal) basis of $\mathbf{R}^n$, that is, $\mathbf{e}_i = (\delta_{i1},\delta_{i2},\ldots,\delta_{in})$ |
| $\lvert\mathbf{v}\rvert$ | = length of vector $\mathbf{v}$ |
| $\mathbf{pq}$ | = the segment from point $\mathbf{p}$ to point $\mathbf{q}$ in $\mathbf{R}^n$, **unless** $\mathbf{p}$ and $\mathbf{q}$ are quaternions, in which case this denotes their product |
| $\lVert\mathbf{pq}\rVert$ | = signed distance from $\mathbf{p}$ to $\mathbf{q}$ |
| $\angle(\mathbf{u},\mathbf{v})$ | = angle between vectors $\mathbf{u}$ and $\mathbf{v}$ |

| | |
|---|---|
| $\angle_s(\mathbf{u},\mathbf{v})$ | = signed angle between vectors **u** and **v** |
| $\mathbf{B}^n(\mathbf{p},r)$ | $= \{\mathbf{q} \in \mathbf{R}^n \mid \lvert\mathbf{pq}\rvert < r\}$ |
| $\mathbf{B}^n(r)$ | $= \mathbf{B}^n(\mathbf{0},r)$ |
| $\mathbf{B}^n$ | $= \mathbf{B}^n(\mathbf{0},1)$ |
| | = the open (n-dimensional) unit disk in $\mathbf{R}^n$ |
| $\mathbf{D}^n(\mathbf{p},r)$ | $= \{\mathbf{q} \in \mathbf{R}^n \mid \lvert\mathbf{pq}\rvert \le r\}$ |
| | = an n-dimensional closed disk |
| $\mathbf{D}^n$ | $= \mathbf{D}^n(\mathbf{0},1)$ |
| | = the closed (n-dimensional) unit disk in $\mathbf{R}^n$ |
| $\mathbf{S}^{n-1}$ | $= \{\mathbf{q} \in \mathbf{R}^n \mid \lvert\mathbf{q}\rvert = 1\}$ |
| | = the (n – 1)-dimensional unit sphere in $\mathbf{R}^n$ |
| $\mathbf{S}^{n-1}_+$ | $= \mathbf{S}^{n-1} \cap \mathbf{R}^n_+$ |
| | = the upper hemisphere |
| $\mathbf{S}^{n-1}_-$ | $= \mathbf{S}^{n-1} \cap \mathbf{R}^n_-$ |
| | = the lower hemisphere |
| $\mathbf{P}^n$ | = n-dimensional projective space |
| $\mathbf{P}^n(k)$ | = n-dimensional projective space over a field k |
| $[\mathbf{L}]$ | = [a,b,c], where **L** is a line in $\mathbf{P}^2$ defined, in homogeneous coordinates, by the equation $aX + bY + cZ = 0$. |

There are natural inclusions: $0 = \mathbf{R}^0 \subset \mathbf{R}^1 \subset \mathbf{R}^2 \subset \ldots$
Similarly for the other spaces above.

The map f: $\mathbf{S}^n \to \mathbf{S}^n$, $f(\mathbf{p}) = -\mathbf{p}$, is called the *antipodal map* of $\mathbf{S}^n$ and **p** and $-\mathbf{p}$ are called *antipodal points*.

| | |
|---|---|
| $\mathbf{X}^k$ | = the k-fold Cartesian product $\underbrace{\mathbf{X} \times \mathbf{X} \times \cdots \times \mathbf{X}}_{k}$ of the set **X** |
| $\mathbf{X} \Delta \mathbf{Y}$ | $= (\mathbf{X} - \mathbf{Y}) \cup (\mathbf{Y} - \mathbf{X})$ (symmetric difference) |
| inf **X** | = infimum or greatest lower bound of the set **X** of real numbers |
| sup **X** | = supremum or least upper bound of the set **X** of real numbers |
| cl(**X**) | = closure of **X** |
| int(**X**) | = interior of **X** |
| aff(**X**) | = affine hull of **X** |
| conv(**X**) | = convex hull of **X** |
| $f(a^+)$ | = right-handed limit of f at a |
| $f(a^-)$ | = left-handed limit of f at a |
| $f^{(d)}(x)$ | = the dth derivative of f |
| $I = I_n$ | = $n \times n$ *identity matrix* that consists of 1s along the diagonal and 0s elsewhere |
| $E_{ij}(c)$ | = $n \times n$ elementary matrix that consists of 1s on the diagonal, the value c in the ijth position, and 0s elsewhere (if i = j, then the ith element on the diagonal is c, not 1) |
| $D(c_1,c_2, \ldots ,c_n)$ | = $n \times n$ diagonal matrix whose ith diagonal entry is $c_i$ and which has 0s elsewhere |
| $A^T$ | = transpose of the matrix A |
| det(A) | = determinant of matrix A |
| tr(A) | = trace of matrix A |

| | |
|---|---|
| $\mathbf{v}^T$ | = the column vector form ($n \times 1$ matrix) of the row vector $\mathbf{v}$ ($1 \times n$ matrix) |
| $\mathbf{GL}(n,k)$ | = the linear group of nonsingular $n \times n$ matrices over $k = \mathbf{R}$ or $\mathbf{C}$ |
| $\mathbf{O}(n)$ | = the group of real orthogonal $n \times n$ matrices |
| $\mathbf{SO}(n)$ | = the group of real special orthogonal $n \times n$ matrices |
| ker h | = kernel of a homomorphism |
| im h | = image of a homomorphism |
| V(f) | = set of zeroes of f (see pages 468, 675, and 676) |
| $<a,b, \ldots>$ | = ideal generated by elements a, b, . . . in a ring |
| $\sqrt{I}$ | = the radical of an ideal I |
| $R(f,g) = R_X(f,g)$ | = the resultant of polynomials f(X) and g(X) |
| k[V] | = ring of polynomial function on $\mathbf{V}$ |
| k(V) | = field of rational functions on $\mathbf{V}$ |
| $tr_k(K)$ | = transcendence degree of field K over k |
| $\bar{z}$ | = the complex conjugate of the complex number z |
| $1_{\mathbf{X}}$ | = the identity map on the set $\mathbf{X}$ |
| $\chi_{\mathbf{A}}$ | = the characteristic function of a set $\mathbf{A}$ as a subset of a given larger set $\mathbf{X}$ |
| $f^{-1}(y)$ | $= \{x \mid f(x) = y\}$ |
| $a \mid b$ | = a divides b |
| Sign(x) | = +1 if $x \geq 0$ and −1 otherwise (returns an **integer**) |
| Sign(σ) | = sign of permutation σ<br>= +1 if σ is an even permutation, −1 if σ is an odd permutation |
| atan2(y,x) | = undefined, if $x = y = 0$,<br>$\pi/2$, if $x = 0$ and $y > 0$,<br>$-\pi/2$, if $x = 0$ and $y < 0$,<br>0, if $y = 0$ and $x > 0$,<br>π, if $y = 0$ and $x < 0$, and<br>θ, where $-\pi < \theta < \pi$, $\tan\theta = y/x$, and θ lies in the same quadrant as (x,y). |

**Note:** atan2(y,x) is closely related to the ordinary arctangent $\tan^{-1}(y/x)$. However, the ordinary arctangent, which is a function of one variable, is not able to keep track of the quadrant in which (x,y) lies, whereas atan2 does. For example,

$$\text{atan2}(-3,-3) = -\frac{3\pi}{4}, \quad \text{but} \quad \text{atan2}(3,3) = \frac{\pi}{4}.$$

| | |
|---|---|
| exp(x) | $= e^x$ |
| $L(\mathbf{V},\mathbf{W})$ | = vector space of linear (see page 873) |
| $L^k(\mathbf{V}_1,\mathbf{V}_2, \ldots ,\mathbf{V}_k;\mathbf{W})$ | = Vector space of multilinear maps (see page 875) |
| $\tau_{\mathbf{M}}$ | = tangent bundle of manifold $\mathbf{M}$ |
| $\nu_{\mathbf{M}}$ | = normal bundle of manifold $\mathbf{M}$ |

| | |
|---|---|
| Vect (**M**) | = vector space of vector fields on the manifold **M** |
| $\boldsymbol{\kappa}_S(s)$ | = signal curvature function of curve in $\mathbf{R}^2$ |
| $\boldsymbol{\kappa}(s)$ | = curvature function |
| $\boldsymbol{\tau}(s)$ | = torsion function |
| $<\boldsymbol{\sigma}>$ | = simplicial complex generated by simplex $\sigma$ |
| $[\boldsymbol{\sigma}]$ | = oriented simplex $\sigma$ |
| $C_q(K)$ | = group of q-chains |
| $B_q(K)$ | = group of q-boundaries |
| $Z_q(K)$ | = group of q-cycles |
| $H_q(K)$ | = q-th homology group |
| $\approx$ | = homeomorphic, isomorphic |
| $\simeq$ | = homotopic, homologous |
| $\simeq_{\mathbf{A}}$ | = homotopic relative to **A** |

Partial derivative notation for a function f:

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \text{ or } f_x, f_y \text{ for } D_1 f, D_2 f, \text{respectively, etc,}$$

$$\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y} \text{ or } f_{xx}, f_{yy} \text{ for } D_{1,1} f, D_{1,2} f, \text{respectively, etc.}$$

$T(\mathbf{A},\mathbf{B}, \ldots) = (\mathbf{A}',\mathbf{B}', \ldots)$ : This means that $T(\mathbf{A}) = \mathbf{A}'$, $T(\mathbf{B}) = \mathbf{B}'$, . . .

**Commutative diagram:** In general, if one has a directed graph where the nodes are sets and the arrows correspond to maps between these sets, then this is said to constitute a commutative diagram if, whenever two directed paths start and end at the same points, the corresponding composition of maps is equal. Commutative diagrams are nice to have and the terminology is useful in many areas of mathematics. As an example, consider the diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ g\downarrow & & \downarrow F \\ C & \xrightarrow[G]{} & D \end{array}$$

If $G(g(a)) = F(f(a))$ for all $a \in A$, then the diagram is said to be *commutative*.

APPENDIX B

# Basic Algebra

## B.1 Number Theoretic Basics

**Definition.** If a and b, $b \neq 0$, are integers and if $a = kb$ for some integer k, then we say that b *divides* a and that b is a *divisor* of a. We write $b|a$.

**Definition.** A positive integer p greater than 1 whose only integer divisors are $\pm 1$ or $\pm p$ is called a *prime number*. Two integers a and b are said to be *relatively prime* if $\pm 1$ are the only common divisors. Given nonzero integers $n_1, n_2, \ldots, n_k$, the *greatest common divisor* of these integers, denoted by $\gcd(n_1,n_2, \ldots ,n_k)$ or simply $(n_1,n_2)$ if $k = 2$, is the largest integer that divides all the $n_i$. The *least common multiple* of these integers, denoted by $\mathrm{lcm}(n_1,n_2, \ldots ,n_k)$, is defined to be the smallest nonnegative integer m so that $n_i$ divides m for all i.

**B.1.1. Theorem.** If a and b are integers that have a greatest common divisor d, then there are integers s and t such that

$$sa + tb = d.$$

**Proof.** This follows from the Euclidean algorithm for integers. See, for example, [Mill58].

**Definition.** Let m be an integer. Two integers a and b are said to *congruent modulo m*, or *mod m*, if m divides $a - b$, equivalently, $a = b + km$ for some k. In that case we shall write

$$a \equiv b \pmod{m}.$$

**Definition.** Let a and m be integers. If $(a,m) = 1$, then a is called a *quadratic residue* modulo m if the congruence

$$x^2 \equiv a \pmod{m}$$

has a solution; otherwise, a is called a *quadratic nonresidue* modulo m.

## B.2 Set Theoretic Basics

**Definition.** A *(binary) relation between sets* ***X*** *and* ***Y*** is a subset **S** of the Cartesian product $\mathbf{X} \times \mathbf{Y}$. If $\mathbf{X} = \mathbf{Y}$, then we call **S** a *relation on* ***X***. The notation x**S**y is often used to denote the fact that $(x,y) \in \mathbf{S}$.

**Definition.** Let **S** be a relation between sets **X** and **Y**. Define subsets dom(**S**) and range(**S**), called the *domain* and *range* of **S**, respectively, by

$$\text{dom}(\mathbf{S}) = \{x \mid x\mathbf{S}y \text{ for some } y \text{ in } \mathbf{Y}\}$$

and

$$\text{range}(\mathbf{S}) = \{y \mid x\mathbf{S}y \text{ for some } x \text{ in } \mathbf{X}\}.$$

The relation **S** is said to be *one-to-one* if $x_1\mathbf{S}y$ and $x_2\mathbf{S}y$ imply that $x_1 = x_2$. The relation **S** is said to be *onto* **Y** if for all y in **Y** there is an x in **X** so that x**S**y. **S** is said to be *well defined* if $x\mathbf{S}y_1$ and $x\mathbf{S}y_2$ imply that $y_1 = y_2$.

**Definition.** Let **S** be a relation on a set **X**. Below are names and definitions of some common properties such a relation may possess:

| | |
|---|---|
| *Reflexive*: | For all $x \in \mathbf{X}$, x**S**x |
| *Symmetric*: | For all $x, y \in \mathbf{X}$, if x**S**y, then y**S**x |
| *Antisymmetric*: | For all $x, y \in \mathbf{X}$, if x**S**y and y**S**x, then $x = y$ |
| *Transitive*: | For all $x, y, z \in \mathbf{X}$, if x**S**y and y**S**z, then x**S**z |

**Definition.** An *equivalence relation* on a set **X** is a reflexive, symmetric, and transitive relation on **X**.

**B.2.1. Example.** It is easy to show that the congruence relation $\equiv$ on the set of integers is an equivalence relation.

The reason that equivalence relations play such an important role in mathematics is that they capture a fundamental concept.

**B.2.2. Theorem.** Let **X** be a nonempty set.

(1) Given a relation **S** on **X** and $x \in \mathbf{X}$, define

$$\mathbf{S}_x = \{y \in \mathbf{X} \mid x\mathbf{S}y\}.$$

If S is an equivalence relation, then each $\mathbf{S}_x$ is nonempty and **X** is the disjoint union of these sets.

(2) Conversely, assume that **X** is the disjoint union of nonempty sets $\mathbf{A}_\alpha$. Define a relation **S** on **X** by the condition that $(x,y) \in \mathbf{S}$ if and only if both x and y belong to $\mathbf{A}_\alpha$ for some $\alpha$. Then **S** is an equivalence relation.

**Proof.** Straightforward.

**Definition.** Let **S** be an equivalence relation on a set **X**. If $x \in \mathbf{X}$, then *the equivalence class of x*, denoted by [x], is defined by

$$[x] = \{y \in \mathbf{X} \mid x\mathbf{S}y\}.$$

The *quotient space of **X** by* **S**, denoted by **X/S**, is defined by

$$\mathbf{X}/\mathbf{S} = \{[x] \mid x \in \mathbf{X}\}.$$

**Definition.** Let **S** be a relation on a set **X**. The equivalence relation on **X** *induced by* **S**, or the *induced equivalence relation*, is defined to be intersection of all equivalence relations on **X** that contain **S**.

One can easily show that the equivalence relation induced by a relation is an equivalence relation and one can think of it as the "smallest" equivalence relation containing **S**.

**Definition.** A well-defined relation **S** between two sets **X** and **Y** is called a *function* or *map* from the domain of **S** to **Y**. A one-to-one function is called an *injective* function or *injection*. An onto function is called a *surjective function* or *surjection*.

We shall use the notation

$$f: \mathbf{X} \to \mathbf{Y}$$

to mean that f is a function between sets **X** and **Y** whose domain is **X** and call f a function from **X** to **Y**. Given such a function, the standard notation for y, given xfy, is f(x) and one typically uses the notation y = f(x) when talking about f. One sometimes calls the range of f the points *traced out* by f.

**Definition.** Given a set **X**, define the *diagonal map*

$$d: \mathbf{X} \to \mathbf{X} \times \mathbf{X} \times \cdots \times \mathbf{X}$$

by

$$d(x) = (x, x, \ldots, x).$$

**Definition.** Given maps $f_i : \mathbf{X}_i \to \mathbf{Y}_i$, define the *product* map

$$f_1 \times f_2 \times \cdots \times f_n : \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_n \to \mathbf{Y}_1 \times \mathbf{Y}_2 \times \cdots \times \mathbf{Y}_n.$$

by

$$(f_1 \times f_2 \times \cdots \times f_n)(x_1, x_2, \ldots, x_n) = (f_1(x_1), f_2(x_2), \ldots, f_n(x_n)).$$

**Definition.** The *graph* of a function $f:\mathbf{X} \to \mathbf{Y}$, denoted by graph(f), is the set that defines it, that is,

$$\text{graph}(f) = \{(x,f(x)) \mid x \in \mathbf{X}\} \subseteq \mathbf{X} \times \mathbf{Y}.$$

**Definition.** Given two functions $f:\mathbf{X} \to \mathbf{Y}$ and $g:\mathbf{Y} \to \mathbf{Z}$, the *composite function*

$$g \circ f: \ \mathbf{X} \to \mathbf{Z}$$

is defined by $(g \circ f)(x) = g(f(x))$, $x \in \mathbf{X}$.

**Definition.** A function $f:\mathbf{X} \to \mathbf{Y}$ that is one-to-one and onto is called *a bijective function or* a *bijection*. In this case, the *inverse* of f, denoted by $f^{-1}$, is the function

$$f^{-1}: \ \mathbf{Y} \to \mathbf{X}$$

defined by the condition that $f^{-1}(y) = x$ if and only if $f(x) = y$.

If the function $f:\mathbf{X} \to \mathbf{Y}$ has an inverse, then it is easy to see that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity functions on $\mathbf{X}$ and $\mathbf{Y}$, respectively. There is a converse to this.

**B.2.3. Theorem.** If $f:\mathbf{X} \to \mathbf{Y}$ and $g:\mathbf{Y} \to \mathbf{X}$ are two functions with the property that $g \circ f$ and $f \circ g$ are the identity maps of $\mathbf{X}$ and $\mathbf{Y}$, respectively, then $g = f^{-1}$.

**Proof.** Straightforward.

The characterization of the inverse of a function in Theorem B.2.3 is often what is used to determine if a function is the inverse of another one.

**Definition.** Let $f:\mathbf{X} \to \mathbf{Y}$ and let $\mathbf{A} \subseteq \mathbf{Y}$. The *inverse image* of $\mathbf{A}$ under f, denoted by $f^{-1}(\mathbf{A})$, is defined by

$$f^{-1}(\mathbf{A}) = \{x \in \mathbf{X} \mid f(x) \in \mathbf{A}\}.$$

For example, if $f(x) = x^2$, then $[3,5] = f^{-1}([9,25])$.

**Definition.** Let $f:\mathbf{X} \to \mathbf{X}$. Any $\mathbf{x}$ in $\mathbf{X}$ with the property that $f(\mathbf{x}) = \mathbf{x}$ is called a *fixed point* of the map f. If $\mathbf{A} \subseteq \mathbf{X}$ and if $f(\mathbf{a}) \in \mathbf{A}$ for all $\mathbf{a} \in \mathbf{A}$, then we say that $\mathbf{A}$ is *invariant* under f or that $\mathbf{A}$ is an *invariant* or *fixed set* for f.

Note that a fixed set is **not** necessarily a set of fixed points. For example, the map $f(x,y) = (x + 1,y)$ has the x-axis as a fixed set but none of the points on the x-axis are fixed points.

**Definition.** A set $\mathbf{X}$ is *countable* if it is finite or there is a bijection between $\mathbf{X}$ and the set of positive integers.

We finish with some definitions of orderings for the elements of a set.

**Definition.** Let $\mathbf{X}$ be a set. A *partial order* on $\mathbf{X}$ is a reflexive, antisymmetric, and transitive relation on $\mathbf{X}$. A *total order* on $\mathbf{X}$ is a partial order $\leq$ with the additional property that for all $x, y \in \mathbf{X}$ either $x \leq y$ or $y \leq x$. A *well-ordering* on $\mathbf{X}$ is a total order $\leq$ with the property that every nonempty subset $\mathbf{A}$ of $\mathbf{X}$ has a *smallest element* in $\mathbf{A}$, that is, there is an $a \in \mathbf{A}$ such that $a \leq a'$ for all $a' \in \mathbf{A}$.

**Notation.** Given a total order $\leq$ on a set $\mathbf{X}$, we shall let $<$ denote the relation on $\mathbf{X}$ defined by $x < y$ if and only if $x \leq y$ and $x \neq y$. Note that the relation $<$ obtained in this way is not a total order because it is not reflexive, but we get back to the total order $\leq$ by simply adding the relations $x \leq x$. Throughout this book, when we have a relation denoted by "$<$" and call it an ordering, we will always assume that it came from some associated total order $\leq$. Expressions such as "the total order $<$" or "the well-ordering $<$" will mean "the total order $\leq$" or "the well-ordering $\leq$", respectively.

The standard example of a total order is the usual $\leq$ relation on numbers, which explains our choice of notation for such a relation. The usual $\leq$ relation is a well-ordering of the set of nonnegative integers. On the other hand, the usual $<$ relation is not a total order on numbers by itself because it is not reflexive. If one desires to have stand-alone definitions for natural orderings like $<$, one can introduce a notion of **strict** partial orders (which basically removes the reflexive property), **strict** total orders, and a corresponding version of well-ordering and smallest element. This is not needed in this book. It should also be noted that there are some slight variations in the definitions for orderings in the literature and readers might have to be careful when they see these terms. However, any differences are technical only and do not change the essential concepts they are trying to capture.

## B.3 Permutations

**Definition.** A *permutation* of a set $\mathbf{X}$ is a bijection $\sigma : \mathbf{X} \to \mathbf{X}$. The set of permutations on $\mathbf{X}$ will be denoted by $S(\mathbf{X})$. Given two permutations $\sigma$ and $\tau$ of $\mathbf{X}$, define their *product* $\sigma \circ \tau$ to be the composition of the two maps, namely,

$$(\sigma \circ \tau)(x) = \sigma(\tau(x)) \quad \text{for } x \in \mathbf{X}.$$

The operation $\circ$ actually makes $S(\mathbf{X})$ into a (noncommutative) group (see the next section for the definition of a group), but this will not play a role in the current discussion.

**Definition.** The set of permutations of $\{1,2, \ldots ,n\}$ together with the operation $\circ$ is called the *symmetric group of degree n* and is denoted by $S_n$.

This section collects a few basic facts about $S_n$.

**Definition.** A permutation $\tau$ in $S_n$ is called a *transposition* if it interchanges two numbers and leaves all others fixed, that is, there are $i, j \in \{1,2, \ldots ,n\}$ with $i \neq j$ such that $\tau(i) = j$, $\tau(j) = i$, and $\tau(k) = k$, for $k \neq i$ or $j$.

**B.3.1. Lemma.** Every permutation $\sigma$ in $S_n$, $n \geq 2$, can be written as a product of transpositions, that is,

$$\sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k,$$

where $\tau_i$ is a transposition.

**Proof.** We use induction on n. The lemma is clearly true if $n = 2$. Note that the identity map is the composition $\tau \circ \tau$ for any transposition $\tau$. Assume that the lemma is true for $n - 1$, $n \geq 3$, and let $\sigma$ be a permutation in $S_n$.

**Case 1.** $\sigma(n) = n$.

In this case, $\sigma$ can be thought of as belonging to $S_{n-1}$. The inductive hypothesis implies that $\sigma$ is a product of transpositions in $S_{n-1}$. Since transpositions of $\{1,2, \ldots ,n - 1\}$ can be thought of as transpositions of $\{1,2, \ldots ,n\}$, we are done.

**Case 2.** $\sigma(n) = i$, $i \neq n$.

Let $\tau_1$ be the transposition in $S_n$ defined by $\tau_1(n) = i$ and $\tau_1(i) = n$. Then

$$(\tau_1 \circ \sigma)(n) = \tau_1(\sigma(n)) = \tau_1(i) = n.$$

By Case 1, $\tau_1 \circ \sigma = \tau_2 \circ \tau_3 \circ \cdots \circ \tau_k$, for some transpositions $\tau_i$. It follows that $\sigma = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_k$, which proves Case 2 and hence the Lemma.

The way in which a permutation can be written as a product of transpositions is not unique, but we have

**B.3.2. Theorem.** Let $\sigma$ be a permutation in $S_n$ and suppose that

$$\begin{aligned} \sigma &= \tau_1 \circ \tau_2 \circ \cdots \circ \tau_s \\ &= \eta_1 \circ \eta_2 \circ \cdots \circ \eta_t, \end{aligned}$$

where $\tau_i$ and $\eta_j$ are transpositions in $S_n$. Then s and t are either both even or both odd.

**Outline of Proof.** If $f: \mathbf{R}^n \to \mathbf{R}$ and $\sigma \in S_n$, then define $f_\sigma: \mathbf{R}^n \to \mathbf{R}$ by

$$f_\sigma(x_1, x_2, \ldots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}). \tag{B.1}$$

Now consider the function $\Delta: \mathbf{R}^n \to \mathbf{R}$ given by

$$\Delta(x_1, x_2, \ldots, x_n) = \prod_{i<j} (x_j - x_i).$$

One can show that $\Delta$ and $\Delta_\sigma$ (as defined by equation (B.1)) satisfies the following two properties:

(1) If $\sigma, \tau \in S_n$, then $\Delta_{\sigma\circ\tau} = (\Delta_\tau)_\sigma$.
(2) For each transposition $\tau$ in $S_n$, $\Delta_\tau = -\Delta$.

Using (1) and (2) we see that for the $\sigma$ in the theorem

$$\begin{aligned}\Delta_\sigma &= \Delta_{\tau_1\circ\tau_2\circ\cdots\circ\tau_s} = (-1)^s\Delta \\ &= \Delta_{\eta_1\circ\eta_2\circ\cdots\circ\eta_t} = (-1)^t\Delta.\end{aligned}$$

In other words, $(-1)^s = (-1)^t$ and the theorem is proved.

Theorem B.3.2 shows that the next definition is well defined.

**Definition.** A permutation $\sigma$ in $S_n$ is said to be *even* if it can be written as a product of an even number of transpositions. Otherwise, $\sigma$ is said to be *odd*.

Clearly, the product of two even permutations is even. Also, $\sigma$ in $S_n$ is even if and only if $\sigma^{-1}$ is even. Therefore, if we define $\sigma \sim \tau$ whenever $\sigma\circ\tau^{-1}$ is even, then $\sim$ is an equivalence relation on $S_n$ and $S_n$ gets partitioned into two equivalence classes by $\sim$, namely, the even and odd permutations.

**Definition.** The *sign* of a permutation $\sigma$, denoted by sign $(\sigma)$, is defined by

$$\begin{aligned}\text{sign}(\sigma) &= +1, \text{ if } \sigma \text{ is even,} \\ &= -1, \text{ if } \sigma \text{ is odd.}\end{aligned}$$

**Definition.** A function $f: \mathbf{X}^d \to \mathbf{Y}$ is said to be *symmetric* if

$$f(x_1, x_2, \ldots, x_d) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(d)})$$

for all $x_i \in \mathbf{X}$ and all permutations $\sigma$ of $\{1,2, \ldots ,d\}$.

## B.4 Groups

The word "group" comes up in several places in this book. The concept is really only essential in Chapters 7 and 8, which involve algebraic topology, but it is useful elsewhere because it does capture some important properties in a **single** word. This section will only survey those definitions and results that are needed in this book. More is not possible. In fact, with the exception of the fundamental group of a topological space which involves free groups and references the concept of the commutator subgroup, all the groups will be abelian. For that reason, other than giving the necessary definitions, most of the discussion and examples will concentrate on abelian groups. The interested reader is directed to books on modern algebra for a more thorough discussion of groups, in particular, the references in the bibliography.

Let G be a set and let $\cdot$ be a binary operation on G, that is, $\cdot$ is a map

$$\cdot : G \times G \to G.$$

As usual in this case, if $g_1$ and $g_2$ belong to G, then we shall write $g_1 \cdot g_2$ instead of the functional form $\cdot\,(g_1,g_2)$.

**Definition.** The pair $(G,\cdot)$ is called a *group* provided that the operation $\cdot$ satisfies:

(1) (Associativity) For all elements $g_1$, $g_2$, and $g_3$ in G,

$$g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3.$$

(2) (Identity) There is an element e in G, called the *identity* of G, such that

$$e \cdot g = g \cdot e = g,$$

for all g in G.

(3) (Inverse) For every g in G, there is an element $g^{-1}$ in G, called the *inverse* of g, such that

$$g \cdot g^{-1} = g^{-1} \cdot g = e.$$

**B.4.1. Example.** The symmetric group of degree n, $S_n$, along with product operation $\circ$, which we discussed in the last section is a finite group with n! elements.

It is not hard to show that the identity and the inverse of each element in a group is unique. Also, it is always the case that

$$\left(g^{-1}\right)^{-1} = g.$$

The simplest group is one that consists only of an identity element.

**Definition.** A group consisting only of an identity element is called a *trivial group*.

**Definition.** A group $(G,\cdot)$ is said to be an *abelian* or *commutative group* if it also satisfies:

(4) (Commutativity) For all $g_1$ and $g_2$ in G, $g_1 \cdot g_2 = g_2 \cdot g_1$.

**B.4.2. Example.** The standard examples of abelian groups are **Z**, **Q**, **R**, and **C** with respect to addition. The sets $\mathbf{Q} - 0$, $\mathbf{R} - 0$, and $\mathbf{C} - 0$ become groups under multiplication. The sets $\mathbf{R}^n$ and $\mathbf{Z}^n$ become groups using the *vector addition*

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

**B.4.3. Example.** For each positive integer n let

$$\mathbf{Z}_n = \{0,1,2,\ldots,n-1\}$$

and define an operation $+_n$ on $\mathbf{Z}_n$ as follows: If $a, b \in \mathbf{Z}_n$, then set

$$a +_n b = d,$$

where

$$a + b = kn + d,$$

and k and d are integers with $0 \le d < n$. It is easy to check that $(\mathbf{Z}_n, +_n)$ (or $\mathbf{Z}_n$ for short) is an abelian group called the *group of integers modulo n* (or *mod n*).

**B.4.4. Example.** Let p be a prime number and let $\mathbf{X} = \{1,2, \ldots ,p-1\}$. Define an operation $\cdot$ on $\mathbf{X}$ as follows: Let $a, b \in \mathbf{X}$. Choose integers k and r so that

$$ab = kp +_n r, \quad \text{where} \quad 0 < r < p.$$

and set

$$a \cdot b = r.$$

It is easy to check that $(\mathbf{X},\cdot)$ is a well-defined abelian multiplicative group. Note that since any integer in $\mathbf{X}$ is relatively prime to p, Theorem B.1.1 implies that it has an inverse with respect to $\cdot$.

**Note.** In the future, when dealing with a group $(G,\cdot)$ for which the operation $\cdot$ is obvious from the context, we shall simply refer to "the group G."

In the case of a general group one typically thinks of the group operation as a multiplication operation and uses "1" as the group identity. The trivial groups consisting of only one element will be denoted by **1**. One also simply writes "gh" for a product of group elements g and h rather than "$g \cdot h$". If g is a group element, then $g^k$ will denote the element

$$\underbrace{g \cdot g \cdot \cdots \cdot g}_{\text{k times}}$$

In the abelian group case, the standard convention is to use additive notation, so that we shall use "+" for the group operation, "0" will be the (additive) identity, and "–g" will denote the inverse of g. The trivial abelian groups consisting of only one element will be denoted by **0**. If $k \in \mathbf{Z}$, then kg will denote the element

$$\underbrace{g + g + \cdots + g}_{\text{k times}}$$

**Definition.** Let $(G,\cdot)$ and $(H,\cdot')$ be groups. We say that $(H,\cdot')$ is a *subgroup* of $(G,\cdot)$ provided that $H \subseteq G$ and $\cdot'$ agrees with $\cdot$, that is, $\cdot' = \cdot \mid (H \times H)$.

**B.4.5. Example.** Two trivial subgroups of any group are the subgroup consisting of only the identity and the whole group itself.

**B.4.6. Example.** For each integer k,

$$k\mathbf{Z} = \{kn \mid n \in \mathbf{Z}\}$$

is a subgroup of **Z** (with respect to addition) and all subgroups of **Z** are of that form.

**B.4.7. Example.** $\mathbf{Z}^n$ is a subgroup of $\mathbf{R}^n$ (with respect to vector addition).

**B.4.8. Example.** The set $\{0,3,6\}$ defines a subgroup of $\mathbf{Z}_9$.

**B.4.9. Example.** The set $\{2n \mid n \in \mathbf{Z}\} \cup \{3\}$ is **not** a subgroup of **Z**.

The next lemma gives a simple criterion for when a subset of a group is a subgroup.

**B.4.10. Lemma.**

(1) A nonempty subset H of a group G is a subgroup (under the operation induced from that of G) if and only if the element $h_1h_2^{-1}$ belongs to H for all $h_1$ and $h_2$ in H.
(2) The intersection of an arbitrary number of subgroups is a subgroup.

**Proof.** Straightforward.

**Definition.** A subgroup H of a group G is said to be a *normal subgroup* of G if for all g in G and all h in H, $ghg^{-1}$ belongs to H.

Note that every subgroup of an abelian group is normal.

**Definition.** Let G and H be groups. A map $f: G \to H$ is called a *homomorphism* if

$$f(g_1g_2) = f(g_1)f(g_2)$$

for all $g_1$ and $g_2$ in G. The homomorphism is said to be an *isomorphism* if it is a bijection. In that case we say that the group G is *isomorphic* to the group H and write $G \approx H$.

It is easy to see that homomorphisms map the identity to the identity and inverses to inverses. In the abelian case this means that $f(0) = 0$ and $f(-g) = -f(g)$ for a homomorphism f. If f is an isomorphism, then so is its inverse $f^{-1}$.

**B.4.11. Example.** Inclusion maps such as $\mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R}$ are clearly homomorphisms.

**B.4.12. Example.** Define a map

$$\pi_n : \mathbf{Z} \to \mathbf{Z}_n$$

as follows: If $k \in \mathbf{Z}$ and $k = an + b$, where a and b are integers and $0 \le b < n$, then let $\pi_n(k) = b$. The map $\pi_n$ is a homomorphism.

Next, let $f: G \to H$ be a homomorphism of groups.

**Definition.** The *kernel* of f, ker(f), and the *image* of f, im(f), are defined by

$$\ker(f) = \{g \in G \mid f(g) = 1\}$$

and

$$\mathrm{im}(f) = \{f(g) \in H \mid g \in G\}.$$

**B.4.13. Lemma.**

(1) The sets ker(f) and im(f) are subgroups of G and H, respectively, with ker(f) a normal subgroup.
(2) The homomorphism f is one-to-one if and only if ker(f) = **1**.

**Proof.** Easy.

Now let H be a subgroup of a group G.

**Definition.** Let $g \in G$. The sets

$$gH = \{gh \mid h \in H\}$$

and

$$Hg = \{hg \mid h \in H\}$$

are called the *left* and *right coset*, respectively, of H in G generated by g. If G is abelian, then the left and right coset are the same set and will be called simply a *coset* in that case.

It can be shown that two left cosets of H in G are either the same sets or they are totally disjoint and the union of all left cosets is G. The same holds for right cosets. In the abelian case, if we define a relation ~ in G by

$$g_1 \sim g_2 \quad \text{if} \quad g_1 - g_2 \in H,$$

then ~ is an equivalence relation and the cosets of H in G are nothing but the equivalence classes of ~.

**B.4.14. Example.** The cosets of {0,3,6} in $\mathbf{Z}_9$ are {0,3,6}, {1,4,7}, and {2,5,8} .

**B.4.15. Example.** The cosets of $\mathbf{Z} = \mathbf{Z}^1$ in $\mathbf{Z}^2$ are the sets $\{(n,k) \mid n \in \mathbf{Z}\}$ for integers k.

Cosets have the same number of elements. One can also show that every left coset is a right coset if the subgroup is normal.

**Definition.** The *factor* or *quotient group* of a group G by a normal subgroup H is defined to be the pair (G/H, ·′), where

$$G/H = \{gH \mid g \in G\}$$

and

$$(g_1H) \cdot' (g_2H) = (g_1 \cdot g_2)H,$$

for all $g_1, g_2 \in G$.

It is straightforward to check that the normality of H implies that $(G/H, \cdot')$ is a well-defined group usually denoted simply by G/H. If H is the trivial subgroup, then one always identifies G/H with G in the natural way.

**B.4.16. Example.** If $G = \mathbf{Z}_9$ and $H = \{0,3,6\}$, then $G/H \approx \mathbf{Z}_3$ .

**B.4.17. Example.** If $G = \mathbf{Z}^2$ and $H = \mathbf{Z}$, then $G/H \approx \mathbf{Z}$ .

**B.4.18. Example.** If $G = \mathbf{Z}$ and $H = \{kn \mid k \in \mathbf{Z}\}$, then $G/H \approx \mathbf{Z}_n$.

**B.4.19. Lemma.** Let G and H be groups. If $f: G \to H$ is a surjective homomorphism, then

$$H \approx G/(\ker f).$$

**Proof.** Let $K = \ker f$. Simply check that the map

$$\begin{aligned} G/K &\to H \\ gK &\to f(g) \end{aligned}$$

is an isomorphism.

**Definition.** A group G is said to be *cyclic* if $G = \{1, g, g^2, \ldots\}$ for some fixed element g in G.

It is easy to see that a cyclic group G is abelian and that the group G in the definition is just $\mathbf{Z}g$ using additive notation. The standard examples of cyclic groups are

$$\mathbf{Z} = \mathbf{Z}1$$

and

$$\mathbf{Z}_n = \mathbf{Z}k, \text{ for } k \text{ in } \mathbf{Z}_n \text{ and } k \text{ relatively prime to } n.$$

It follows from the next lemma that there are no others.

**B.4.20. Lemma.** A cyclic group G is isomorphic to either $\mathbf{Z}$ or $\mathbf{Z}_n$ for some positive integer n.

**Proof.** Let $G = \mathbf{Z}g$ and define a homomorphism

$$\varphi : \mathbf{Z} \to G$$

by

$$\varphi(k) = kg.$$

If $\ker(\varphi) = \mathbf{0}$, then $G \approx \mathbf{Z}$; otherwise,

$$\ker(\varphi) = \{kn \mid k \in \mathbf{Z}\}$$

for some $n \in \mathbf{Z}$ and $G \approx \mathbf{Z}_n$.

**Definition.** Let G be a group and let g be an element of G. Define the *order* of g, o(g), to be the smallest of the integers $k > 0$, such that $g^k = 1$, if such integers exist; otherwise, define the *order* of g to be $\infty$.

**B.4.21. Example.** Let $G = \mathbf{Z}_6 = \{0,1,2,3,4,5\}$. Then $o(1) = 6 = o(5)$, $o(2) = 3 = o(4)$, $o(3) = 2$, and $o(0) = 1$.

**Definition.** The number of elements in a group G is called the *order* of G and is denoted by o(G). If G has only a finite number of elements, then G is called a *group of finite order*, or simply a *finite group*.

Note that two finite groups of the same order need not be isomorphic.

**B.4.22. Theorem.** (Lagrange) Let G be a finite group.

(1) If H is a subgroup of G, then $o(H) \mid o(G)$. In fact, if n is the number of cosets of H in G, then $o(G) = n\, o(H)$. The number n is called the *index* of H in G.
(2) If $g \in G$, then $o(g) \mid o(G)$.

**Proof.** Part (1) follows from the fact that G is a union of disjoint cosets of H, each of which has the same number of elements. To prove (2), one only has to observe that $o(g) = o(\mathbf{Z}g)$ and apply (1).

**Definition.** A *commutator* of a group G is any element in G of the form $ghg^{-1}h^{-1}$, where $g, h \in G$.

**B.4.23. Theorem.** The set of finite products of commutators of G is a normal subgroup H called the *derived* or *commutator subgroup* of G. The factor group G/H is abelian and called the *abelianization* of G.

**Proof.** See [Dean66].

Finally, we want to define a free group. Basically, a free group is a group in which no "nontrivial" identities hold between elements. An example of a trivial identity is $gg^{-1} = 1$, which holds for all elements g in a group. A nontrivial identity would be something like $g_1g_2{}^5g_3 = 1$, if that were to hold for all elements $g_i$. Although the idea

of a free group is fairly simple, the definition is surprisingly complicated. We shall sketch it only.

Let S be any set of elements called symbols. Think of the elements of S as the letters in an alphabet S. A *reduced word* in S will refer to any formal string of the form

$$s_1^{n_1} s_2^{n_2} \cdots s_k^{n_k}$$

where $s_i \in S$, $n_i \in \mathbf{Z}$, and we have made all possible elementary contractions. An elementary contraction involves replacing occurrences of the form $s^n s^m$ by $s^{n+m}$, occurrences of $s^0$ by 1, and occurrences of $1 \cdot s$ for $s \cdot 1$ by s. For example, if $S = \{a,b,c\}$, then $a^2bc^{-7}$ is a reduced word but $a^2a^{-2}bc$ is not since it can be reduced to bc. Let F(S) denote the set of reduces words in S. If $u, v \in F(S)$, define $u \cdot v$ to be the element of F(S) obtained by concatenating u and v and then reducing the result as much as possible.

**B.4.24. Lemma.** $(F(S),\cdot)$ is a group.

**Definition.** $(F(S),\cdot)$ is called the *free group generated by S*. The elements of S are called *generators* of the group F(S).

A key property of free groups is that they satisfy a universality property with respect to being able to extend maps of S into a group to a homomorphism of F(S).

**B.4.25. Theorem.** Let G be any group and let $h: S \to G$ be any map. Then h extends to a unique homomorphism $H: F(S) \to G$.

**Proof.** See [Fral67]. To put it another way, each h lifts to a unique map H so that we have a commutative diagram

$$\begin{tikzcd} F(S) \arrow[dr, "H"] & \\ S \arrow[u, "\text{inclusion map } i", "\cup"'] \arrow[r, "h"'] & G \end{tikzcd}$$

This *universal factorization property* is what actually defines the free group F(s) in the sense that if K is any other group containing S and satisfying this property, then K is isomorphic to F(S).

**Definition.** Let $G_1, G_2, \ldots,$ and $G_n$ be groups. The *direct product* of the $G_i$, denoted by $G_1 \times G_2 \times \cdots \times G_n$, is the group $(G_1 \times G_2 \times \cdots \times G_n,\cdot)$, where the operation $\cdot$ is defined by

$$(g_1, g_2, \ldots, g_n) \cdot (g_1', g_2', \ldots, g_n') = (g_1 \cdot g_1', g_2 \cdot g_2', \ldots, g_n \cdot g_n').$$

(The operations on the right are the operations from the groups $G_i$.)

It is easy to check that $(G_1 \times G_2 \times \cdots \times G_n,\cdot)$ is a group.

## B.5 Abelian Groups

In this section we concentrate on abelian groups and we shall use the standard additive notation.

**Definition.** Let G be an abelian group and let $g_1, g_2, \ldots, g_k \in G$. Then

$$\mathbf{Z}g_1 + \mathbf{Z}g_2 + \cdots + \mathbf{Z}g_k = \{n_1g_1 + n_2g_2 + \cdots + n_kg_k \mid n_i \in \mathbf{Z}\}$$

is called the *subgroup* of G *generated by* the $g_1, g_2, \ldots,$ and $g_k$.

It is easy to show that $\mathbf{Z}g_1 + \mathbf{Z}g_2 + \cdots + \mathbf{Z}g_k$ is in fact a subgroup of G and also that it is the intersection of all subgroups of G which contain the elements $g_1, g_2, \ldots,$ and $g_k$.

**B.5.1. Lemma.** For any abelian group G the elements of finite order form a unique subgroup T(G) called the *torsion subgroup* of G.

**Proof.** The proof is clear since $o(0) = 1$, $o(-g) = o(g)$, and $o(g + h) \mid o(g)o(h)$ for all $g, h \in G$.

**Definition.** An abelian group G is said to be *torsion-free* if it has no element of finite order other than 0, that is, $T(G) = \mathbf{0}$.

Clearly, G/T(G) is a torsion-free group for every abelian group G.

**Definition.** An abelian group G is said to be *finitely generated* if

$$G = Zg_1 + Zg_2 + \cdots + Zg_n$$

for some $g_1, g_2, \ldots, g_n \in G$. In that case, the $g_i$ are called *generators* for G.

**B.5.2. Example.** All cyclic groups are finitely generated.

**B.5.3. Example.** The group $\mathbf{Z}^2$ is not cyclic, but it is finitely generated since $\mathbf{Z}^2 = \mathbf{Z}(1,0) + \mathbf{Z}(0,1)$. A similar statement holds for $\mathbf{Z}^n$, $n \geq 2$ .

**B.5.4. Example.** The groups **Q** and **R** are not finitely generated.

**B.5.5. Lemma.** Subgroups of finitely generated abelian groups are finitely generated.

**Proof.** See [Dean66].

**Definition.** Let $G_1, G_2, \ldots,$ and $G_n$ be abelian groups. The *external direct sum* of the $G_i$, denoted by $G_1 \otimes G_2 \otimes \cdots \otimes G_n$, is the group $(G_1 \times G_2 \times \cdots \times G_n, +)$, where the operation + is defined by

$$(g_1, g_2, \ldots, g_n) + (g_1', g_2', \ldots, g_n') = (g_1 + g_1', g_2 + g_2', \ldots, g_n + g_n').$$

It is easy to check that $(G_1 \times G_2 \times \cdots \times G_n, +)$ is an abelian group. Except for the different name and notation that is used in the abelian case, the external direct sum is just the direct product defined at the end of the last section.

**B.5.6. Example.** The groups $\mathbf{Z}^n$ and $\mathbf{R}^n$ are the external direct sums of n copies of $\mathbf{Z}$ and $\mathbf{R}$, respectively.

**Definition.** Let $G_1, G_2, \ldots,$ and $G_n$ be subgroups of an abelian group G. Define the *sum* of the groups $G_i$, denoted by $G_1 + G_2 + \cdots + G_n$, by

$$G_1 + G_2 + \cdots + G_n = \{g_1 + g_2 + \cdots + g_n \mid g_i \in G_i\}.$$

It is easy to show that $G_1 + G_2 + \cdots + G_n$ is a subgroup of G.

**Definition.** If $G_1, G_2, \ldots,$ and $G_n$ are subgroups of a given abelian group G and if each element g in G can be written uniquely in the form $g_1 + g_2 + \cdots + g_n$ with $g_i$ in $G_i$, then G is called the *internal direct sum* of the groups $G_i$ and we write $G = G_1 \oplus G_2 \oplus \cdots \oplus G_n$.

Because it is easily shown that $G_1 \oplus G_2 \oplus \cdots \oplus G_n$ is isomorphic to $G_1 \otimes G_2 \otimes \cdots \otimes G_n$, one usually does not distinguish between internal and external direct sums and uses the same symbol $\oplus$ to denote either direct sum. The context decides which is being used. One also uses the summation notation, so that the expressions

$$G_1 \oplus G_2 \oplus \cdots \oplus G_n, \quad G_1 \otimes G_2 \otimes \cdots \otimes G_n, \quad \text{and} \quad \bigoplus_{i=1}^{n} G_i$$

will all refer to the same object.

**B.5.7. Theorem.** (The Fundamental Theorem of Finitely Generated Abelian Groups) Let G be a finitely generated abelian group. Then

$$G \approx \underbrace{\mathbf{Z} \oplus \mathbf{Z} \oplus \cdots \oplus \mathbf{Z}}_{r} \oplus \mathbf{Z}_{n_1} \oplus \mathbf{Z}_{n_2} \oplus \cdots \oplus \mathbf{Z}_{n_t},$$

for some r, $t \geq 0$, where $1 < n_i \in \mathbf{Z}$ and $n_i \mid n_{i+1}$ if $t > 0$. The integer r is called the *rank* of G and is denoted by rank(G). The integers $n_1, n_2, \ldots,$ and $n_t$ are called the *torsion coefficients* of G. Both the rank and the torsion coefficients are uniquely determined by G.

**Proof.** See [Dean66].

Note that the example

$$\mathbf{Z}_2 \oplus \mathbf{Z}_3 \approx \mathbf{Z}_6$$

shows that the condition that $n_i$ divides $n_{i+1}$ is important for the uniqueness part of Theorem B.5.7.

**B.5.8. Theorem.** Assume that G, H, and K are finitely generated abelian groups.

(1) If $\varphi: K \to G$ and $\psi: G \to H$ are homomorphisms satisfying

(a) $\varphi$ is injective,
(b) $\mathrm{im}(\varphi) = \ker(\psi)$, and
(c) $\psi$ is surjective,

then

$$\mathrm{rank}(G) = \mathrm{rank}(K) + \mathrm{rank}(H).$$

(2) The function rank is "additive", that is,

$$\mathrm{rank}(H \oplus K) = \mathrm{rank}(H) + \mathrm{rank}(K).$$

Before proving Theorem B.5.8, we introduce some more notation.

**Definition.** The subset $\{g_1, g_2, \ldots, g_n\}$ of an abelian group G is said to form a *basis* for G provided that

(1) $G = \mathbf{Z}g_1 + \mathbf{Z}g_2 + \cdots + \mathbf{Z}g_n$, and
(2) if $k_1g_1 + k_2g_2 + \cdots + k_ng_n = 0$, for $k_i \in \mathbf{Z}$, then $k_ig_i = 0$ for all i.

It is easy to show that $\{g_1, g_2, \ldots, g_n\}$ is a basis for the group G if and only if

$$G = \mathbf{Z}g_1 \oplus \mathbf{Z}g_2 \oplus \cdots \oplus \mathbf{Z}g_n.$$

Thus, by Theorem B.5.7 each finitely generated group has a basis.

**Definition.** Any abelian group which is isomorphic to $G_1 \oplus G_2 \oplus \cdots \oplus G_n$, where each $G_i$ is isomorphic to $\mathbf{Z}$, is called a *free abelian group*.

It follows easily from Theorem B.5.7 and the definitions that the following conditions on a finitely generated abelian group are equivalent:

(1) G is free.
(2) G is torsion-free.
(3) G has a basis consisting of elements of infinite order.

**Proof of Theorem B.5.8.** To prove part (1), assume without loss of generality that all three groups G, H, and K are torsion-free. Let $\{h_1, h_2, \ldots, h_s\}$ and $\{k_1, k_2, \ldots, k_t\}$ be bases for H and K, respectively, where $s = \mathrm{rank}(H)$ and $t = \mathrm{rank}(K)$. Choose elements $g_1, g_2, \ldots,$ and $g_s$ in G such that $\psi(g_i) = h_i$. It is now easy to show that

$$\{\varphi(k_1), \varphi(k_2), \ldots, \varphi(k_t), g_1, g_2, \ldots, g_s\}$$

forms a basis for G. In other words,

$$\text{rank}(G) = t + s = \text{rank}(K) + \text{rank}(H).$$

Part (2) follows from (1) by letting $G = H \oplus K$ and letting $\varphi$ and $\psi$ be the natural inclusion and projection map, respectively, and the theorem is proved.

**B.5.9. Theorem.** Let G be a free abelian group with basis $\{g_1, g_2, \ldots, g_n\}$. If H is any group and $h_1, h_2, \ldots, h_n \in H$, then there exists a unique homomorphism $\varphi: G \to H$ such that $\varphi(g_i) = h_i$.

**Proof.** Show that any element g of G has a unique representation of the form

$$g = k_1 g_1 + k_2 g_2 + \cdots + k_n g_n, \quad k_i \in \mathbf{Z},$$

and define

$$\varphi(g) = k_1 h_1 + k_2 h_2 + \cdots + k_n h_n.$$

Free abelian groups behave very similarly to free groups. Compare Theorem B.5.9 with Theorem B.4.25. There is also an analogous commutative diagram to describe what is going on and we again have a *universal factorization property*. Alternatively, Theorem B.5.9 can be paraphrased by saying that, given a free group G and any other group H, in order to define a homomophism from G to H it suffices to define it on a basis of G, because any map from a basis of G to H extends uniquely to a homomorphism of G. The freeness of G is important because consider

$$G = \mathbf{Z}_3 = \{0,1,2\}$$

and define

$$\varphi(1) = 1 \in \mathbf{Z}.$$

The map $\varphi$ does **not** extend to a homomorphism $\varphi: \mathbf{Z}_3 \to \mathbf{Z}$. Of course, although 1 is a basis for $\mathbf{Z}_3$, the group $\mathbf{Z}_3$ is not a free group and so there is no contradiction.

Next, let G and H be groups.

**Definition.** Let

$$\text{hom}(G,H) = \{h \mid h : G \to H \text{ is a homomorphism}\}$$

and if $h_i \in \text{hom}(G,H)$, then define

$$h_1 + h_2 : G \to H$$

by

$$(h_1 + h_2)(g) = h_1(g) + h_2(g) \quad \text{for } g \in G.$$

**B.5.10. Lemma.** (hom (G,H),+) is an abelian group.

**Proof.** Straightforward.

**B.5.11. Lemma.** Let G be any group that is isomorphic to **Z**. If $h \in$ hom (G,G), then there is a unique integer k such that h(g) = kg for all g in G.

**Proof.** Since G is isomorphic to **Z**, there is some $g_0$ in G so that $G = \mathbf{Z}g_0$. It follows that $h(g_0) = kg_0$ for some integer k, because $\{g_0\}$ is a basis for G. The fact that $o(g_0) = \infty$ implies that the integer k is unique. Now let g be any element of G. Again, there is some integer t with $g = tg_0$. Thus,

$$h(g) = h(tg_0) = th(g_0) = t(kg_0) = k(tg_0) = kg,$$

and the lemma is proved.

Lemma B.5.11 implies that if $G \approx \mathbf{Z}$, then hom $(G,G) = \mathbf{Z}1_G \approx \mathbf{Z}$.

## B.6 Rings

**Definition.** A *ring* is a triple $(R,+,\cdot)$ where R is a set and + and · are two binary operations on R, called addition and multiplication, respectively, satisfying the following:

(1) (R,+) is an abelian group.
(2) The multiplication · is associative.
(3) For all a, b, c ∈ R we have
 (a) (*left distributativity*) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
 (b) (*right distributativity*) $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$

Two standard examples of rings are **Z** and $\mathbf{Z}_n$.

**Definition.** A ring in which the multiplication is commutative is called a *commutative ring*. A ring with a multiplicative identity is called a *ring with unity*. An element of a ring with unity is called a *unit* if it has a multiplicative inverse in R.

**Definition.** Let $(R,+,\cdot)$ be a ring. If A is a subset of R and if $(A,+,\cdot)$ is a ring, then $(A,+,\cdot)$ is called a *subring* of R.

**Note.** From now on, like in the case of groups, we shall not explicitly mention the operations + and · for a ring$(R,+,\cdot)$ and simply refer to "the ring R." Products "r · s" will be abbreviated to "rs."

**Definition.** A subset I of a ring R with the property that

(1) I is an additive subgroup of R (equivalently, $a - b \in I$ for all $a, b \in I$), and
(2) ra, ar ∈ I for all r ∈ R and a ∈ I,

is called an *ideal* of R.

An ideal is a subring. Note that if the ring R is commutative, then we can replace condition (2) in the definition of an ideal by

(2′) $ra \in I$ for all $r \in R$ and $a \in I$.

**Definition.** Let R and R′ be rings. A map $h: R \to R'$ is said to be a (ring) *homomorphism* if

$$h(a+b) = h(a) + h(b) \quad \text{and}$$
$$h(ab) = h(a)h(b),$$

for all $a, b \in R$. If h is a bijection, then h is called a (ring) *isomorphism*.

**Definition.** Let $h: R \to R'$ be a homomorphism between rings. The *kernel* of h, ker h, and the *image* of h, im(h), are defined by

$$\ker h = \{r \in R \mid h(r) = 0\}$$

and

$$\mathrm{im}(h) = \{h(r) \mid r \in R\}.$$

**B.6.1. Theorem.**

(1) The image of a ring homomorphism is a subring.
(2) The kernel of a homomorphism is an ideal.
(3) A ring homomorphism is an isomorphism if and only if its kernel is **0**.

**Proof.** See [Fral67].

**Definition.** Let I be an ideal in a ring $(R,+,\cdot)$. The *factor* or *quotient ring* of the ring R by the ideal I, denoted by $(R/I,+,\cdot)$, is defined as follows: The additive group $(R/I,+)$ is just the quotient group of $(R,+)$ by the subgroup $(I,+)$. The operation $\cdot$ is defined by

$$(r+I)\cdot(s+I) = rs + I.$$

Using the definition of an ideal it is easy to check that the quotient ring of a ring R by an ideal I is in fact a ring and that the map

$$R \to R/I$$

that sends an element into its coset is a surjective ring homomorphism with kernel I.

In analogy with the integers one can define a notion of congruence.

**Definition.** Let I be an ideal in a commutative ring R. Let $a, b \in R$. We say that a and b are *congruent modulo I*, or *mod I*, and write

$$a \equiv b \pmod{I},$$

if $a - b \in I$.

It is easy to show that $\equiv$ is an equivalence relation on R and the equivalence classes are just the cosets of the additive subgroup of R.

**Definition.** Let $r_1, r_2, \ldots,$ and $r_k$ be elements of a commutative ring R. then

$$< r_1, r_2, \ldots, r_k > = \{a_1 r_1 + a_2 r_2 + \cdots + a_k r_k \mid a_i \in R\}$$

is called the *ideal generated by the $r_i$*.

It is easy to show that $<r_1, r_2, \ldots, r_k>$ is an ideal.

**Definition.** If r is an element of a commutative ring R, then $<r>$ is called the *principal ideal generated by r* in R.

**Definition.** An ideal I in a ring R, $I \neq R$, is called a *maximal* ideal if, whenever J is an ideal of R with $I \subseteq J \subseteq R$, then either $J = I$ or $J = R$.

**Definition.** A commutative ring R, $I \neq R$, with identity is said to be an *integral domain* if $ab = 0$ implies that either $a = 0$ or $b = 0$.

**Definition.** An integral domain is called a *principal ideal domain* (PID) if all of its ideals are principal ideals.

The ring of integers is a good example of a principal ideal domain.

**Definition.** Let a and b be elements of a commutative ring R. We say that b *divides* a, denoted by $b|a$, if b is nonzero and $a = bc$, for some c in R. One calls b a *factor* or *divisor* of a in this case.

**Definition.** Let R be a commutative ring with unity. Two elements a and b in R are said to be *associates* if $a = ub$, where u is a unit.

**Definition.** Let R be an integral domain. An element a in R is said to be *irreducible* if

(1) a is not 0,
(2) a is not a unit,
(3) for all b in R, if $b|a$, then b is a unit or b is an associate of a.

**Definition.** An integral domain is called a *unique factorization domain* (UFD) provided that

(1) Every nonzero element r is either a unit or a product of irreducible elements.
(2) If

$$r = p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m,$$

where the $p_i$ and $q_j$ are irreducible elements, then n = m and upon renumbering we may assume that $p_i$ and $q_i$ differ by a unit factor.

The ring of integers is the standard example of a UFD.

**B.6.2. Theorem.** Every PID is a UFD.

**Proof.** See [Fral67].

**Definition.** Let I and J be ideals in a commutative ring R. Define the *product* of I and J, denoted by $I \cdot J$ or IJ, by

$$I \cdot J = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n \mid a_i \in I \text{ and } b_i \in J \text{ for all } i\}.$$

Define the *sum* of I and J, denoted by I + J, by

$$I + J = \{a + b \mid a \in I, b \in J\}.$$

It is easy to show that $I \cdot J$, the set of finite linear combinations of products of elements from I and J, is actually an ideal. One can rephrase the definition of an ideal in a **commutative** ring R as follows:

$$\text{An ideal is an additive subgroup I with } R \cdot I \subseteq I.$$

One can also show that the sum of ideals is an ideal. In fact, if I, J, and K are ideals, then the following properties hold:

(1) Associativity: $I \cdot (J \cdot K) = (I \cdot J) \cdot K$
(2) Commutativity: $I \cdot J = J \cdot I$
(3) Distributivity: $I \cdot (J + K) = I \cdot J + I \cdot K$
$(I + J) \cdot K = I \cdot K + J \cdot K$

What this means is that using the product and sum on ideals along with standard set operations, one can define a factorization theory for ideals.

**Definition.** Let I and J be ideals in a commutative ring R. We say that I is a *divisor* of J and that J is a *multiple* of I if $I \supseteq J$.

For example, <5> is a divisor of <15> and <15> is a multiple of <5> in **Z**.

**Definition.** An ideal I in a commutative ring R is called a *prime* ideal if, whenever ab lies in I for some elements a and b in R, then either a or b belongs to I. Equivalently, the ideal I is prime if $ab \equiv 0 \pmod{I}$ implies that either $a \equiv 0 \pmod{I}$ or $b \equiv 0 \pmod{I}$.

For example, <7> is a prime ideal in **Z**.

**B.6.3. Theorem.** Let R be a commutative ring with unity element and let I be an ideal in R. Then I is a prime ideal if and only if R/I is an integral domain.

**Proof.** See [Mill58].

**Definition.** Let I be an ideal in a commutative ring R. The *radical* of I, denoted by $\sqrt{I}$, is defined by

$$\sqrt{I} = \{a \in R \mid a^n \in I \text{ for some } n > 0\}.$$

An ideal I is called a *radical ideal* if $I = \sqrt{I}$.

**B.6.4. Proposition.** Let I be an ideal in a commutative ring R.

(1) $\sqrt{I}$ is an ideal that divides I.
(2) $\sqrt{I}$ is a radical ideal.
(3) If I is prime, then I is a radical ideal.

**Proof.** Easy.

The next definition generalizes the notion of a prime power element.

**Definition.** An ideal I in a commutative ring R is called a *primary* ideal if whenever $ab \in I$ for some b that does not lie in I, then $a \in \sqrt{I}$.

We can rephrase the definition to say that every zero divisor of a primary ideal I is in its radical, that is, $ab \equiv 0 \pmod{I}$ and $b \not\equiv 0 \pmod{I}$ implies that $a \equiv 0 \pmod{\sqrt{I}}$.

**B.6.5. Proposition.** The radical of a primary ideal in a commutative ring is a prime ideal.

**Proof.** Easy.

For example, note that the ideal $<q>$ is primary in $\mathbf{Z}$ if and only if $q = p^n$ where p is a prime.

Now the equation

$$n = p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$$

implies that

$$<n> = <p_1^{k_1}> \cap <p_2^{k_2}> \cap \cdots \cap <p_m^{k_m}>.$$

The natural question is whether such a factorization of ideals holds in general. The answer is yes, provided that the ring satisfies certain chain conditions.

**Definition.** A commutative ring R is said to satisfy the *ascending chain condition* if for every sequence of ideals $I_i$ satisfying

$$I_1 \subseteq I_2 \subseteq \cdots$$

there is an n, so that $i > n$ implies that $I_i = I_{i+1}$.

**Definition.** A commutative ring R is said to be a *Noetherian ring* provided that it satisfies the ascending chain condition.

**B.6.6. Theorem.** A commutative ring R is Noetherian if and only if every ideal is finitely generated.

**Proof.** See [Jaco66] or [ZarS60], Volume I.

**Definition.** An ideal I in a commutative ring is said to be *reducible* if I can be expressed as the intersection of two ideals $I_1$ and $I_2$, $I = I_1 \cap I_2$, where $I_j \neq I$. Otherwise, I is said to be *irreducible*.

A prime ideal is irreducible, but a primary ideal need not be. An irreducible ideal is not necessarily prime. For example, $<p^k>$ is irreducible in **Z** but not prime if $k > 1$. On the other hand,

**B.6.7. Lemma.** Every irreducible ideal in a Noetherian ring is primary.

**Proof.** See [Jaco66] or [ZarS60], Volume I.

**Definition.** An intersection

$$I = I_1 \cap I_2 \cap \cdots \cap I_n$$

of ideals in a ring is said to be an *irredundant intersection* if I is a proper subset of

$$I_1 \cap I_2 \cap \cdots \cap I_{i-1} \cap I_{i+1} \cap \cdots \cap I_n$$

for $i = 1, 2, \ldots, n$.

**B.6.8. Theorem.** Every ideal in a Noetherian ring is a finite irredundant intersection of primary ideals and the prime ideals associated to this factorization are unique.

**Proof.** See [Jaco66] or [ZarS60], Volume I.

## B.7 Polynomial Rings

One of the important examples of rings are polynomial rings.

**Definition.** Let A be a subring and S a subset of a ring R. The *polynomial ring over A generated by S*, denoted A[S], is defined by

$$A[S] = \cap \{B \mid B \text{ is a subring of } R \text{ and } A, S \subseteq B\}.$$

It is easy to show that the intersection of subrings is a subring, so that A[S] is a subring. Furthermore, if A is a subring of a commutative ring R with unity and if u is any element of R, then it is easy to show that

$$A[u] = \{a_0 + a_1u + a_2u^2 + \cdots + a_nu^n \mid a_i \in R \quad \text{and} \quad n \geq 0\}.$$

This justifies calling A[u] a polynomial ring. We would like to define the polynomial ring R[X] in an "indeterminate" (or "variable") X, where "indeterminate" basically means that X is transcendental over R in the following sense:

**Definition.** Let A be a subring of a commutative ring R with unity and let $u \in R$. We say that u is *transcendental* over A if

$$a_0 + a_1u + a_2u^2 + \cdots + a_nu^n = 0$$

with $a_i \in R$ implies that $a_i = 0$ for all i. If u is not transcendental, then u is said to be *algebraic* over A.

The definition of R[X] boils down to constructing an object with the desired properties. Because polynomials are a special case of formal power series, we shall deal with both simultaneously so as not to have to duplicate basically the same definitions later on.

**Definition.** Let R be a ring. A *formal power series* over R is an infinite sequence $(a_0,a_1,a_2,\ldots)$ where the $a_i$ are elements of R. If all but a finite number of the $a_i$ are zero, then the sequence is called a *polynomial over R*. The power series for which all the $a_i$ are zero is called the *zero power series* or *zero polynomial* and will be denoted by 0. In the case of a nonzero polynomial, the largest index i for which $a_i$ is nonzero is called the *degree* of the polynomial. The zero polynomial is said to have degree 0.

Let $f = (a_0,a_1,a_2,\ldots)$ and $g = (b_0,b_1,b_2,\ldots)$ be formal power series over a ring R. Define an addition + and a multiplication · of formal power series as follows:

$$f + g = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots)$$

and

$$f \cdot g = (c_0, c_1, c_2, \ldots), \quad \text{where} \quad c_k = \sum_{i=0}^{k} a_i b_{k-i}.$$

Let X be some symbol (or *indeterminate*). We shall identify the polynomial

$$(\underbrace{0,0,\ldots,0}_{i-1},a,0,\ldots)$$

with the formal expression "$aX^i$." The terms "$aX^0$" and "$aX^1$" will be abbreviated to "a" and "aX," respectively. With this identification a formal power series

$$f = (a_0, a_1, a_2, \ldots)$$

can then be expressed in the more usual form as

$$f = f(X) = a_0 + a_1X + a_2X^2 + \cdots = \sum_{i=0}^{\infty} a_iX^i.$$

In the case of a polynomial of degree n, one writes f in the form

$$f = f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n = \sum_{i=0}^{n} a_iX^i. \tag{B.2}$$

**Definition.** The term $a_nX^n$ in expression (B.2) is called the *leading term* of f and $a_n$ is called the *leading coefficient* of f and are denoted by lt(f) and lc(f), respectively.

**B.7.1. Theorem.** If R[[X]] is the set of formal power series over R, then (R[[X]],+,·) is a ring. If R is commutative, then so is R[[X]]. If R is an integral domain, then so is R[[X]]. If R[X] ⊂ R[[X]] is the set of polynomials over R, then (R[X],+,·) is a subring of R[[X]].

**Proof.** Easy.

**Definition.** R[[X]] is called the *formal power series ring over R* (in the indeterminate X). R[X] is called the *polynomial ring over R* (in the indeterminate X).

Formal power series are the algebraic analog of power series in calculus. The difference is that we do not worry about convergence here and there is no need to define a topology for the ring.

**B.7.2. Theorem.** If R is a subring of a ring Q and if u is an element of Q that is transcendental over R, then R[X] is isomorphic to R[u].

**Proof.** The theorem follows from a *universal factorization property* satisfied by R[X] with respect to ring homomorphisms, namely,

> Let R′ be a subring of Q′ and let u′ ∈ Q′. Any ring homomorphism $h: R \to R'$ extends to a unique ring homomorphism $H: R[X] \to R'[u']$, which maps X to u′.

**B.7.3. Theorem.** If R is a UFD, then both R[[X]] and R[X] are UFDs. The only irreducible element in R[[X]], up to unit, is X.

**Proof.** For R[X] see [Dean66], for example. For R[[X]] see [Seid68].

Let R be a ring and let $X_1, X_2, \ldots$ be symbols (or *indeterminates*). Define rings $R^{((n))}$ and $R^{(n)}$ recursively by

$$R^{((0))} = R$$
$$R^{((n))} = R^{((n-1))}[[X_n]], \quad \text{for} \quad n > 0,$$

and

$$R^{(0)} = R$$
$$R^{(n)} = R^{(n-1)}[X_n], \quad \text{for} \quad n > 0.$$

**Definition.** $R^{((n))}$ is called the *formal power series ring over R in indeterminates $X_1, X_2, \ldots,$ and $X_n$* and is denoted by $R[[X_1,X_2, \ldots ,X_n]]$. $R^{(n)}$ is called the *polynomial ring over R in indeterminates $X_1, X_2, \ldots,$ and $X_n$* and is denoted by $R[X_1,X_2, \ldots ,X_n]$.

Note that with our notation, $R[[X_1,X_2, \ldots ,X_n]] = R[[X_1,X_2, \ldots ,X_{n-1}]] [[X_n]]$ and similarly for the polynomial ring. From this, Theorem B.7.3, and induction on n we get

**B.7.4. Corollary.** If R is a UFD, then both $R[[X_1,X_2, \ldots ,X_n]]$ and $R[X_1,X_2, \ldots ,X_n]$ are UFDs.

Now, every polynomial $f = f(X_1,X_2, \ldots ,X_n) \in R[X_1,X_2, \ldots ,X_n]$ can be written as a finite sum in the form

$$\sum a_{r_1 r_2 \cdots r_n} X_1^{r_1} X_2^{r_2} \cdots X_n^{r_n}, \quad \text{where} \quad a_{r_1 r_2 \cdots r_n} \in R. \tag{B.3}$$

**Definition.** An expression of the form

$$aX_1^{r_1} X_2^{r_2} \cdots X_n^{r_n},$$

where a is a nonzero element of R, is called a *monomial* of *total degree* d, where

$$d = r_1 + r_2 + \cdots + r_n.$$

The element a is called the *coefficient* of the monomial. If a = 1, we call the expression a *power product*.

**Definition.** Let $f \in R[X_1,X_2, \ldots ,X_n]$. The *total degree* or simply *degree* of the polynomial f is the largest total degree of all the monomials appearing in f. We say that f is a *quadratic*, *cubic*, . . . *polynomial* if its degree is 2, 3, . . . , respectively. The polynomial f is said to be *homogeneous of degree d* or simply *homogeneous* if each monomial that appears in it has total degree d.

Note that any formal power series f in $R[[X_1,X_2, \ldots ,X_n]]$ can be written uniquely in the form

$$f = f_0 + f_1 + f_2 + \cdots$$

where each $f_i$ is a homogeneous polynomial of degree i.

**Definition.** If $f_i$ is not the zero polynomial, then $f_i$ is called the ith *homogeneous component* of f.

Every polynomial f can be written uniquely in the form

$$f = f_0 + f_1 + \cdots + f_d$$

where each $f_i$ is a homogeneous polynomial of degree i and d is the total degree of f.

**Definition.** If f is nonzero, then the smallest integer i, so that $f_i \neq 0$, is called the *order* of f and is denoted by ord(f). The order of 0, ord(0), is defined to be $\infty$.

Finally, note that if $f \in R[[X_1, X_2, \ldots, X_n]]$, then for each i, if we let

$$R_i = R[[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n]],$$

then $f \in R_i[[X_i]]$. The same holds for the polynomial rings.

**Definition.** Let

$$f = \sum a_{r_1 r_2 \cdots r_n} X_1^{r_1} X_2^{r_2} \cdots X_n^{r_n}, \quad \text{where} \quad a_{r_1 r_2 \cdots r_n} \in R$$

be a polynomial in $R[X_1, X_2, \ldots, X_n]$. The map

$$\pi_f : R^n \to R$$
$$(c_1, c_2, \ldots, c_n) \to \sum a_{r_1 r_2 \cdots r_n} c_1^{r_1} c_2^{r_2} \cdots c_n^{r_n}$$

is called the *evaluation map* associated to f and $\pi_f(c_1, c_2, \ldots, c_n)$ will be denoted by $f(c_1, c_2, \ldots, c_n)$. If

$$f(c_1, c_2, \ldots, c_n) = 0,$$

then $(c_1, c_2, \ldots, c_n)$ is called a *zero* or *root* of f.

An evaluation map (other than at 0) does not make sense for formal power series without a notion of limit; however, substitution or composition does as long as the series we are substituting has no constant term. A precise definition of the formal power series that is the composition of two power series is actually somewhat technical (see [ZarS60], Volume II or [Walk50]) and we shall not take the time to do this here. Intuitively,

**Definition.** If $f(X), g(X) \in R[[X]]$, then the *composition* of f and g, denoted by $(f \circ g)(X)$, is the power series where X in f(X) is replaced by g(X) and we collect all the coefficients of all the same powers of X in the result. The power series $f \circ g$ is also referred to as the power series obtained from f by *substitution* of g into f.

We cannot allow g to have a constant term, otherwise the composition would potentially have a constant term that would be an infinite sum of elements of R.

**B.7.5. Theorem.** If $f, g \in R[[X]]$, where $\text{ord}(g) > 0$, then $f \circ g$ is well defined. Furthermore,

(1) If $fg \neq 0$, then $\text{ord}(f \circ g) = \text{ord}(f)\,\text{ord}(g)$.
(2) If $h \in R[[X]]$ and $\text{ord}(h) > 0$, then $f \circ (g \circ h) = (f \circ g) \circ h$.

**Proof.** See [Walk50].

**B.7.6. Proposition.** A polynomial $f \in R[X_1,X_2, \ldots ,X_n]$ of degree k is homogeneous of degree k if and only if

$$f(tX_1,tX_2,\ldots,tX_n) = t^k f(X_1,X_2,\ldots,X_n). \tag{B.4}$$

**Proof.** The only nontrivial part is showing that if f satisfies equation (B.4), then it is homogeneous of degree k, but this reduces to the easy case n = 1 by setting $X_2, X_3, \ldots, X_n$ to $X_1$.

**Definition.** A polynomial $f \in R[X_1,X_2, \ldots ,X_n]$ is said to be *symmetric* if

$$f(X_1,X_2,\ldots,X_n) = f(X_{\sigma(1)},X_{\sigma(2)},\ldots,X_{\sigma(n)})$$

for all permutations $\sigma$ of $\{1,2, \ldots ,n\}$.

Consider the equation

$$(Z-X_1)(Z-X_2)\cdots(Z-X_n) = \sigma_0 Z^n - \sigma_1 Z^{n-1} - \sigma_2 Z^{n-2} - \cdots - (-1)^n \sigma_n, \tag{B.5}$$

where the left-hand side is a polynomial in an indeterminate Z over the ring $R[X_1,X_2, \ldots ,X_n]$ and the right-hand side is its expansion, which defines polynomials $\sigma_i = \sigma_i(X_1,X_2, \ldots ,X_n)$. Because the left-hand side of equation (B.5) is unchanged by permuting the $X_i$'s, the polynomials $\sigma_i$ are symmetric.

**Definition.** The polynomial $\sigma_i$ is called the *ith elementary symmetric polynomial* in the variables (indeterminates) $X_1, X_2, \ldots, X_n$.

For example, if n = 3, then

$$\begin{aligned}
&\sigma_1(X_1,X_2,X_3) = X_1 + X_2 + X_3,\\
&\sigma_2(X_1,X_2,X_3) = X_1X_2 + X_1X_3 + X_2X_3, \quad \text{and}\\
&\sigma_3(X_1,X_2,X_3) = X_1X_2X_3.
\end{aligned}$$

The elementary symmetric polynomials form a basis for all symmetric polynomials.

**B.7.7. Theorem.** (The Fundamental Theorem of Symmetric Polynomials) If $f(X_1,X_2, \ldots ,X_n)$ is a symmetric polynomial over a ring R with unity, then there exists a unique polynomial $F(X_1,X_2, \ldots ,X_n)$ over R, so that

$$f(X_1,X_2,\ldots,X_n) = F(\sigma_1,\sigma_2,\ldots,\sigma_n),$$

where $\sigma_i$ is the ith elementary polynomial in $X_1, X_2, \ldots, X_n$.

**Proof.** See [Jaco66].

One can define a formal derivative for power series and polynomials.

**Definition.** Let

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_iX^i + \cdots$$

be a formal power series (or polynomial) over a ring R with unity. Define the *derivative of f*, denoted by $f'(X)$ or $df/dX$, by

$$f'(X) = a_1 + 2a_2X + 3a_3X^2 + \cdots + ia_iX^{i-1} + \cdots.$$

If $f(X_1,X_2, \ldots ,X_n)$ is a multivariate formal power series (or polynomial), define the *ith partial derivative of f*, denoted by $\partial f/\partial X_i$, to be the derivative of f thought of as a formal power series (or polynomial) in $X_i$.

It is easy to see that this definition of derivative or partial derivative of a polynomial matches the corresponding definition that one encounters in calculus if one thinks of the polynomials as functions. It also satisfies the same properties. The main point is that there is no need to introduce the notion of limits and the purely algebraic aspects of the derivative turn out to have important applications in algebra, in particular, algebraic geometry.

**B.7.8. Theorem.** (Euler) If $f(X_1,X_2, \ldots ,X_n)$ is a homogeneous polynomial of degree d in the variables $X_i$, then

$$\sum_{i=1}^{n} X_i \frac{\partial}{\partial X_i} f(X_1, X_2, \ldots, X_n) = df(X_1, X_2, \ldots, X_n).$$

**Proof.** Clearly, it suffices to prove the result in the case where f is a monomial

$$X_1^{d_1} X_2^{d_2} \cdots X_n^{d_n}, \; d_1 + d_2 + \cdots + d_n = d.$$

But in that case,

$$X_i \frac{\partial}{\partial X_i} f(X_1, X_2, \ldots, X_n) = d_i f(X_1, X_2, \ldots, X_n),$$

from which the theorem easily follows.

**B.7.9. Theorem.** (Hilbert Basis Theorem) If R is a Noetherian ring, then so is R[X].

**Proof.** See [Jaco66].

**B.7.10. Corollary.** If R is a Noetherian ring, then so is $R[X_1,X_2, \ldots ,X_n]$.

**Proof.** This uses Theorem B.7.9 and induction.

**B.7.11. Theorem.** If R is a Noetherian ring, then so is $R[[X_1,X_2, \ldots ,X_n]]$.

**Proof.** See [ZarS60], Volume II.

# B.8 Fields

**Definition.** Let R be a ring with unity. If every nonzero element of R is a unit, then R is called a *skew field* or *division ring*. A *field* is a commutative division ring. A subring of a field that is a field is called a *subfield*.

**B.8.1. Example.** **Q**, **R**, and $\mathbf{Z}_p$, where p is prime, are all fields.

It is easy to see that the intersection of fields is again a field.

**Definition.** Let K be a field. The intersection k of all of its nonzero subfields is called the *prime field of K*. If k = K, then K is called a *prime field*.

The prime field of a field can also be described as the "smallest" subfield of a field.

**B.8.2. Theorem.** A prime field is isomorphic to either **Q** or $\mathbf{Z}_p$, where p is prime.

**Proof.** See [Mill58].

**Definition.** Let K be a field and k its prime subfield. If k is isomorphic to $\mathbf{Z}_p$, where p is prime, then K is said to have *characteristic* p. Otherwise, k is isomorphic to **Q** and K is said to have *characteristic* 0.

Every integral domain D can be imbedded in a field. The construction generalizes the way that one gets the rational numbers from the integers. Let

$$Q^* = \{(a,b) \mid a,b \in D \text{ and } b \neq 0\}.$$

Define an equivalence relation ~ on Q* as follows:

$$(a,b) \sim (c,d) \text{ if and only if } ad = cd.$$

It is easy to check that ~ is an equivalence relation. Let Q denote the equivalence classes of Q* with respect to ~. Define two operations + and · on Q:

$$[a,b]+[c,d]=[ad+bc,bd]$$
$$[a,b]\cdot[c,d]=[ac,bd].$$

**B.8.3. Theorem.** (Q,+, ·) is a well-defined field.

**Proof.** See [Mill58].

**Definition.** The field (Q,+, ·) in Theorem B.8.3 is called the *quotient field* of the integral domain D.

**B.8.4. Example.** The field **Q** of rational numbers is the quotient field of the integers **Z**.

**B.8.5. Theorem.** Let R be a commutative ring with unity element and let I be an ideal in R. Then I is a maximal ideal if and only if R/I is a field.

**Proof.** See [Mill58].

**B.8.6. Corollary.** Every maximal ideal in a commutative ring with unity element is prime.

**Proof.** This follows from Theorems B.6.3 and B.8.5.

The next theorem and its proof is the basic division algorithm for polynomials.

**B.8.7. Theorem.** Let k be a field. If $f(X), g(X) \in k[X]$ and $g(X) \neq 0$, then there exist unique polynomials $q(X), r(X) \in k[X]$ such that

(1) $f = q\,g + r$, and
(2) $r = 0$ or degree $r <$ degree $g$.

**Proof.** See [Dean66].

**B.8.8. Theorem.** Let k be a field. If $f(X) \in k[X]$, then for any $c \in k$ there exists a unique polynomial $q(X) \in k[X]$ such that

$$f(X) = (X - c)q(X) + f(c).$$

**Proof.** See [Dean66].

An easy consequence of Theorem B.8.8 is

**B.8.9. Corollary.** Let k be a field. If $f(X) \in k[X]$, then c is a zero of f if and only if $X - c$ divides f.

Let k be a field and $f(X) \in k[X]$. If c is a root of f(X), then it follows from Corollary B.8.9 that we can express f(X) in the form

$$f(X) = (X - c)^n g(X),$$

where c is not a root of g(X).

**Definition.** The integer n is called the *multiplicity* of the root c of f(X). If $n = 1$, then one calls c a *simple root*. If $n > 1$, then one calls c a *multiple root*. More generally, if g(X) is a factor of f(X) and if n is the largest integer with the property that $g(X)^n$ divides f(X), then n is called the *multiplicity* of the factor g(X). We call g(X) a *multiple factor* if $n > 1$ .

**B.8.10. Theorem.** Let k be a field. A nonconstant polynomial $f(X) \in k[X]$ has a multiple factor $g(X)$ if and only if $g(X)$ is also a factor of $f'(X)$.

**Proof.** Let $f(X) = g(X)^n h(X)$. The product rule of the derivative easily leads to the result.

**B.8.11. Theorem.** If k is a field, then k[X] is a PID (and hence a UFD).

**Proof.** See [Mill58].

Only the polynomial ring in **one** variable over a field is a PID. To see that $k[X_1,X_2, \ldots ,X_n]$ is not a PID if $n > 1$, one simply needs to convince oneself that the ideal in k[X,Y] consisting of all polynomials whose constant term is 0 is not a principal ideal.

**Definition.** Let $f_1, f_2, \ldots , f_k \in k[X]$. A *greatest common divisor* of the $f_i$, denoted by $\gcd(f_1,f_2, \ldots ,f_k)$, is a largest degree polynomial $g \in k[X]$ that divides each of the $f_i$. A *least common multiple* of the $f_i$, denoted by $\mathrm{lcm}(f_1,f_2, \ldots ,f_k)$, is a smallest degree polynomial $h \in k[X]$ with the property that each $f_i$ divides h.

**B.8.12. Theorem.** Let $f_1, f_2, \ldots , f_k \in k[X]$.

(1) The polynomials $\gcd(f_1,f_2, \ldots ,f_k)$ and $\mathrm{lcm}(f_1,f_2, \ldots ,f_k)$ are unique up to scalar multiple.
(2) $\langle f_1,f_2, \ldots ,f_k\rangle = \langle \gcd(f_1,f_2, \ldots ,f_k)\rangle$.

**Proof.** This is easy. The key fact is that k[X] is a UFD.

**B.8.13. Theorem.** If k is a field, then both $k[[X_1,X_2, \ldots ,X_n]]$ and $k[X_1,X_2, \ldots ,X_n]$ are Noetherian.

**Proof.** First note that k satisfies the ascending chain condition since it has only two ideals, namely, 0 and itself. Now use Corollary B.7.10 and Theorem B7.11.

**Definition.** Let K be a field. If k is a subfield of K, then K is called an *extension* (field) of k.

**Definition.** Let $K_1$ and $K_2$ be extension fields of a field k. An isomorphism $\sigma: K_1 \to K_2$ is called an *isomorphism over k* if $\sigma$ is the identity on k. If $K = K_1 = K_2$, then $\sigma$ is called an *automorphism of K over k*.

**Definition.** Let k be a subfield of a field K. If $A \subseteq K$, then k(A) will denote the *smallest subfield of K containing k and A*, more precisely,

$$k(A) = \cap \{F \mid F \text{ is a subfield of K which contains k and A}\}.$$

If A consists of a single element a, then we shall usually write k(a) instead of k(A).

It is easy to show that the intersection of fields is a field and so k(A) is a well-defined subfield.

**Definition.** If K is an extension of a field k and if K = k(a) for some element a, then K is called a *simple* extension of k and the element a is called a *primitive element* of K over k.

**Definition.** If k is a field, then k(X) will denote the *field of quotients*, or *quotient field*, of k[X]. More generally, $k(X_1,X_2,\ldots,X_n)$ will denote the field of quotients of $k[X_1,X_2,\ldots,X_n]$.

There are natural inclusions

$$k \subset k(X_1) \subset k(X_1,X_2) \subset \cdots \subset k(X_1,X_2,\ldots,X_n).$$

**B.8.14. Theorem.** Let k be a subfield of a field K. If $a \in K$, then either

$$k(a) \approx k(X)$$

or a is a root of an irreducible polynomial f(X) in k[X] and

$$k(a) \approx k[X]/\langle f(X)\rangle.$$

**Proof.** See [Mill58]. Note that $\langle f(X)\rangle$ is a maximal ideal and so $k[X]/\langle f(X)\rangle$ is a field by Theorem B.8.5.

**B.8.15. Theorem.** Let K and K′ be subfields of fields E and E′, respectively, and let $\varphi: K \to K'$ be an isomorphism. Let

$$p(X) = a_0 + a_1X + \cdots + a_nX^n \in K[X]$$

and

$$p'(X) = \varphi(a_0) + \varphi(a_1)X + \cdots + \varphi(a_n)X^n \in K'[X].$$

Assume that p(X) is irreducible. If $c \in E$ and $c' \in E'$ are roots of the polynomials p(X) and p′(X), respectively, then the isomorphism φ extends to a unique isomorphism $\varphi': K(c) \to K'(c')$ such that $\varphi'(c) = c'$.

**Proof.** See [Dean66].

## B.9 The Complex Numbers

This section will simply define the field of complex numbers. Complex numbers play an essential role in many areas of mathematics. Appendix E will discuss some of their nontrivial properties, especially as they relate to analysis.

Identify $(a,b) \in \mathbf{R}^2$ with the formal expression $a + \mathbf{i}b$. ($0 + \mathbf{i}b$ is abbreviated to $\mathbf{i}b$ and $\mathbf{i}\cdot 1$, to $\mathbf{i}$.) Using this notation define arithmetic operations + and · on $\mathbf{R}^2$ as follows:

$$(a+\mathbf{i}b)+(c+\mathbf{i}d)=(a+c)+\mathbf{i}(b+d)$$
$$(a+\mathbf{i}b)\cdot(c+\mathbf{i}d)=(ac-bd)+\mathbf{i}(ad+bc)$$

**B.9.1. Theorem.** $(\mathbf{R}^2,+,\cdot)$ is a field denoted by $\mathbf{C}$ which contains $\mathbf{R}$ as a subfield under the identification of a with $(a,0)$.

**Proof.** Straightforward.

Thought of as field elements, the elements of $\mathbf{R}^2$ are called *complex numbers*. It is easy to check that we have the well-known identity

$$\mathbf{i}^2 = -1,$$

in other words, $-1$ has a square root in $\mathbf{C}$.

**Definition.** If $z = a + \mathbf{i}b \in \mathbf{C}$, then a is called the *real part* of z and b is called the *imaginary part* of z. Define functions Re(z) and Im(z) by

$$\mathrm{Re}(z)=a \quad \text{and} \quad \mathrm{Im}(z)=b.$$

The *complex conjugate* of z, $\bar{z}$, and the *modulus* of z, $|z|$, are defined by

$$\bar{z}=a-\mathbf{i}b \quad \text{and} \quad |z|=\sqrt{a^2+b^2}.$$

Here are two simple facts that are easily checked:

(1) The modulus function is multiplicative, that is, $|z_1z_2| = |z_1||z_2|$ .
(2) If $z = a + \mathbf{i}b$, then we have the identity

$$\frac{1}{z}=\frac{a}{a^2+b^2}-\mathbf{i}\frac{b}{a^2+b^2}=\frac{1}{|z|}\bar{z}$$

## B.10 Vector Spaces

Vector spaces are discussed in more detail in Appendix C. We only define them here and list those basic properties that are needed in the section on field extensions.

**Definition.** A *vector space over a field k* is a triple $(\mathbf{V},+,\cdot)$ consisting of a set $\mathbf{V}$ of objects called *vectors* together with two operations + and · called *vector addition* and

*scalar multiplication*, respectively, so that $(\mathbf{V},+)$ is an abelian group, and for each $a \in k$ and $\mathbf{u} \in \mathbf{V}$, $a\cdot\mathbf{v} \in \mathbf{V}$. Furthermore, the following identities hold for each $a, b \in k$ and $\mathbf{u}, \mathbf{v} \in \mathbf{V}$:

(1) (distributivity) $a\cdot(\mathbf{u} + \mathbf{v}) = a\cdot\mathbf{u} + a\cdot\mathbf{v}$
(2) (distributivity) $(a + b)\cdot\mathbf{u} = a\cdot\mathbf{u} + b\cdot\mathbf{u}$
(3) (associativity) $(ab)\cdot\mathbf{u} = a\cdot(b\cdot\mathbf{u})$
(4) (identity) $1\cdot\mathbf{u} = \mathbf{u}$

For simplicity, the operator $\cdot$ is usually suppressed and one writes $a\mathbf{u}$ instead of $a\cdot\mathbf{u}$. Also, if the field and the operations are clear from the context, the vector space $(\mathbf{V},+,\cdot)$ will be referred to simply as the vector space $\mathbf{V}$. It is easy to show that

$$-\mathbf{u} = (-1)\mathbf{u}$$

and it is useful to define a subtraction operators for vectors by setting

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}).$$

*N-dimensional Euclidean space* $\mathbf{R}^n$, or *n-space*, is the most well-known example of a vector space because it is more than just a set of points and admits a well-known vector space structure over $\mathbf{R}$. Namely, let $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbf{R}^n$, $c \in \mathbf{R}$, and define

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, \ldots, u_n + v_n)$$
$$c\mathbf{u} = (cu_1, \ldots, cu_n).$$

One thinks of elements of $\mathbf{R}^n$ as either "points" or "vectors," depending on the context. More generally, it is easy to see that if k is a field, then $k^n$ is a vector space over k.

Function spaces are another important class of vector spaces. Let k be a field and let $\mathbf{A}$ be a subset of $k^n$. Then the set of functions from $\mathbf{A}$ to k is a vector space over k by *pointwise addition and scalar multiplication*. More precisely, if $f, g: \mathbf{A} \to k$ and $c \in k$, define

$$f + g, cf : \mathbf{A} \to k$$

by

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (cf)(x) = c(f(x)).$$

**Definition.** A *subspace* of a vector space $\mathbf{V}$ is a subset of $\mathbf{V}$ with the property that it, together with the induced operations from $\mathbf{V}$, is itself a vector space.

For example, under the natural inclusions, the vector spaces $\mathbf{R}^m$, $m \le n$, are all subspaces of $\mathbf{R}^n$.

**Definition.** Let **V** be a vector space over a field k. Given a **nonempty** set of vectors $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ in **V**, we define the *span of the set*, span(S), or the *span of the vectors* in S, $\text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, to be the set of all vectors that are *linear combinations* of these vectors, that is,

$$\text{span}(S) = \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \{c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n \mid c_i \in k\}.$$

We say that the set S *spans* **X** and the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ *span* a subspace **X** if

$$\mathbf{X} = \text{span}(S) = \text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n).$$

It is convenient to define $\text{span}(\phi) = \{\mathbf{0}\}$ .

It is easy to check that $\text{span}(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is a vector subspace of **V**.

**Definition.** Let **V** be a vector space over a field k. A **nonempty** set of vectors $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ in **V**, is said to be *linearly dependent* if

$$c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \mathbf{0}$$

for some field elements $c_1, c_2, \ldots, c_k$ not all of which are zero. The set S and the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are said to be *linearly independent* if they are not linearly dependent. It is convenient to define the empty set to be a linearly independent set of vectors.

In other words, vectors are linearly independent if no nontrivial linear combination of them adds up to the zero vector. Two linearly dependent vectors are often called *collinear*. Two **nonzero** vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are collinear if and only if they are multiples of each other, that is, $\text{span}(\mathbf{u}_1) = \text{span}(\mathbf{u}_2)$.

**Definition.** Let **X** be a subspace of a vector space. A set of vectors S is said to be a *basis* for **X** if it is a linearly independent set that spans **X**.

**Note.** The definitions of linearly independent/dependent, span, and basis above dealt only with finite collections of vectors to make the definitions clearer and will apply to most of our vector spaces. On a few occasions we may have to deal with "infinite dimensional" vector spaces and it is therefore necessary to indicate what changes have to be made to accommodate those. All one has to do is be a little more careful about what constitutes a linear combination of vectors. Given an arbitrary, possibly infinite, set of vectors $\{\mathbf{v}_\alpha\}_{\alpha \in I}$ in a vector space, define a *linear combination* of those vectors to be a sum

$$\sum_{\alpha \in I} c_\alpha \mathbf{v}_\alpha$$

where all but a finite number of the $c_\alpha$ are zero. With this concept of linear combination, the definitions above and the next theorem will apply to **all** vector spaces.

**B.10.1. Theorem.** Every vector subspace of a vector space has a basis and the number of vectors in a basis is uniquely determined by the subspace. Every set of linearly independent vectors in a vector space can be extended to a basis.

**Proof.** See [Dean66].

Theorem B.10.1 justifies the following definition:

**Definition.** The *dimension* of a vector space **V**, denoted by dim **V**, is defined to be the number of vectors in a basis for it. A m-dimensional subspace of an n-dimensional vector space is said to have *codimension* n–m.

**Note.** With our definitions above, the vector space that consists only of the zero vector has dimension 0 and the empty set is a basis for it. This may sound a little strange, but it gives us a nice uniform terminology without which some results would get a little more complicated to state.

**Definition.** Let **V** and **W** be vector spaces over a field k. A map $T:\mathbf{V} \to \mathbf{W}$ is called a *linear transformation* if T satisfies

$$T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w})$$

for all $\mathbf{v} \in \mathbf{V}$, $\mathbf{w} \in \mathbf{W}$, and $a, b \in k$ .

**B.10.2. Theorem.** The inverse of a linear transformation, if it exists, is a linear transformation and so is the composite of linear transformations.

**Proof.** Easy.

**Definition.** A linear transformation is said to be *nonsingular* if it has an inverse; otherwise, it is said to be *singular*. A nonsingular linear transformation is called a vector space *isomorphism*.

**Definition.** Let **V** and **W** be vector spaces over a field k. If $T:\mathbf{V} \to \mathbf{W}$ is a linear transformation, then define the *kernel* of T, ker (T), and the *image* of T, im (T), by

$$\ker(T) = \{\mathbf{v} \in \mathbf{V} \mid T(\mathbf{v}) = 0\}$$

and

$$\operatorname{im}(T) = [T(\mathbf{v}) \mid \mathbf{v} \in \mathbf{V}] \subseteq \mathbf{W}.$$

**B.10.3. Theorem.** If $T:\mathbf{V} \to \mathbf{W}$ is a linear transformation between vector spaces **V** and **W**, then the kernel and image of T are vector subspaces of **V** and **W**, respectively. Furthermore,

$$\dim \mathbf{V} = \dim \operatorname{im}(T) + \dim \ker(T).$$

**Proof.** Easy. See [John67].

Let **W** be a subspace of a vector space **V** over a field k. It is easy to check that the quotient group **V**/**W** becomes a vector space over k by defining

$$c(\mathbf{v} + \mathbf{W}) = c\mathbf{v} + \mathbf{W}$$

for each coset $\mathbf{v} + \mathbf{W}$ in $\mathbf{V}/\mathbf{W}$ and $c \in k$.

**Definition.** The vector space $\mathbf{V}/\mathbf{W}$ is called the *quotient vector space* of $\mathbf{V}$ by $\mathbf{W}$.

It is easy to show that the natural projection $\mathbf{V} \to \mathbf{V}/\mathbf{W}$ that sends a vector in $\mathbf{V}$ to the coset that it determines is a well-defined surjective linear transformation whose kernel is $\mathbf{W}$.

**B.10.4. Theorem.** Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces over a field k. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis for $\mathbf{V}$ and let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ be **any** vectors in $\mathbf{W}$. There exists a unique linear transformation $T: \mathbf{V} \to \mathbf{W}$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$, $i = 1, 2, \ldots, n$.

**Proof.** Easy.

Let k be a field and S a set. Define F(S) to be the set of "formal" sums

$$\sum_{s \in S} r_s s,$$

where $r_s \in k$ and all but a **finite** number of the $r_s$ are 0. There is an obvious addition and scalar multiplication making F(S) into a vector space. A more precise definition of F(S) would define it to be the set of functions $\varphi: S \to k$ that vanish on all but a finite number of elements in S. We leave it to the reader to fill in the technical details.

**Definition.** The vector space F(S) is called the *free vector space over k with basis S*.

**B.10.5. Theorem.** Let S be a set and $\mathbf{V}$ a vector space. If $t: S \to \mathbf{V}$ is any map, then there is a unique **linear** map $T: F(S) \to V$ with $T(s) = t(s)$ for all $s \in S$.

**Proof.** Easy. Like with free groups, the best way to think about this lifting of t to a map T is via a commutative diagram:

$$\begin{tikzcd}
F(S) \arrow[dr, "T"] & \\
S \arrow[u, "\text{inclusion map } i", "\cup"'] \arrow[r, "t"'] & V
\end{tikzcd}$$

We have another *universal factorization property* and this is what actually defines F(S) up to isomorphism.

## B.11 Extension Fields

This section describes some fundamental properties of extension fields.

**Definition.** Let k be a subfield of a field K. An element $a \in K$ is said to be *transcendental over k* if $k(a) \approx k(X)$. It is said to be *algebraic over k* if it is the root of a polynomial $f(X) \in k[X]$. If it is the root of an irreducible polynomial of degree n, then it is said to be *algebraic over k of degree n*.

This definition of transcendental and algebraic agrees with the definition given for rings in Section B.7.

**B.11.1. Theorem.** Let k be a subfield of a field K. If an element $a \in K$ is algebraic over k, then a is the root of a unique irreducible polynomial with leading coefficient equal to 1 called the *minimum polynomial* of a over k.

**Proof.** See [Jaco64].

**Definition.** An extension K of a field k is called an *algebraic* extension of k if every element of K is algebraic over k. If K is not an algebraic extension of k, then it is called a *transcendental* extension.

**Definition.** An extension K of a field k is called a *finite extension of k of degree n* if K is a finite dimensional vector space over k of dimension n. Otherwise, K is called an *infinite extension* of k of degree $\infty$. The degree of the extension is denoted by $[K:k]$.

**B.11.2. Theorem.** If K is a finite extension of a field k, then K is an algebraic extension of k.

**Proof.** See [Mill58].

**B.11.3. Theorem.** If K is a finite extension of a field k, then $K = k(\theta_1, \theta_2, \ldots, \theta_n)$, where the $\theta_i$ are algebraic over k.

**Proof.** See [Mill58].

**B.11.4. Theorem.** (Theorem of the Primitive Element) If k is a field of characteristic 0 and if $\theta_1$, $\theta_2, \ldots,$ and $\theta_n$ belong to some extension field of k and are algebraic over k, then

$$k(\theta_1, \theta_2, \ldots, \theta_n) = k(\theta)$$

for some element $\theta$ (which is algebraic over k). In other words, every finite extension of a field of characteristic 0 is simple.

**Proof.** See [Mill58].

**B.11.5. Theorem.** Let $\theta$ be a transcendental element over a field k and let K be a field such that $k \subset K \subseteq k(\theta)$ with $k \neq K$. Then there exists an element $\sigma$ in K which is transcendental over k and $K = k(\sigma)$.

**Proof.** See [Walk50].

**Definition.** Let $f(X) \in k[X]$ be a polynomial of positive degree. An extension field K of k is called a *splitting field* of f(X) if

(1) f(X) is a product of linear factors in K{X}, that is,

$$f(X) = a(X - \theta_1)(X - \theta_2) \cdots (X - \theta_n)$$

with $a \in k$, $\theta_i \in K$, and

(2) $K = k(\theta_1, \theta_2, \ldots, \theta_n)$.

**B.11.6. Theorem.** Every polynomial over a field k of positive degree has a splitting field. Any two splitting fields of the polynomial are isomorphic over k.

**Proof.** See [Jaco64].

**Definition.** A polynomial $f(X) \in k[X]$ is said to be *separable* if it is a product of irreducible polynomials each of which has only simple roots in a splitting field of f(X).

**B.11.7. Theorem.** Every polynomial over a field of characteristic 0 is separable.

**Proof.** See [Jaco64].

**Definition.** Let K be an extension of a field k. An element of K is said to be *separable* over k if it is algebraic over k and its minimum polynomial is separable. K is a *separable extension* of k if every element of K is separable over k.

**Definition.** Let K be an extension of a field k. The elements $a_1, a_2, \ldots, a_n$ in K are said to be *algebraically dependent* over k if there exists a nonzero polynomial $f(X_1, X_2, \ldots, X_n)$ in $k[X_1, X_2, \ldots, X_n]$ and $f(a_1, a_2, \ldots, a_n) = 0$. Otherwise, the elements are said to be *algebraically independent*. An infinite set of elements of K is said to be *algebraically dependent* over k if each of its finite subsets is algebraically dependent. Otherwise it is said to be *algebraically independent*.

Alternatively, the elements $a_1, a_2, \ldots, a_n$ are algebraically independent over k if the map

$$k[X_1, X_2, \ldots, X_n] \to k[a_1, a_2, \ldots, a_n]$$
$$f(X_1, X_2, \ldots, X_n) \to f(a_1, a_2, \ldots, a_n)$$

is an isomorphism over k.

**Definition.** Let K be an extension of a field k. A maximal set of elements in K that are algebraically independent over k is called a *transcendence basis* for K over k. The number of elements in a transcendence basis is called the *transcendence degree* of K over k and is denoted by $tr_k(K)$.

The notion of transcendence degree is justified by the following fact:

**B.11.8. Theorem.** Any two transcendence bases of K over k have the same cardinal number of elements.

**Proof.** See [ZarS60].

**Definition.** A field k is said to be *algebraically closed* if every nonconstant polynomial f(X) in k[X] has a zero in k.

The complex numbers **C** are the standard example of an algebraically closed field. See Corollary E.5.2 for a proof.

**Definition.** Let K be an extension of a field k. K is called an *algebraic closure* of k if

(1) K is algebraic over k and
(2) K is algebraically closed.

**B.11.9. Theorem.**

(1) Every field k has an algebraic closure.
(2) Any two algebraic closures of k are isomorphic over k.

**Proof.** See [Jaco64].

**B.11.10. Theorem.** If k is an algebraically closed field, then every polynomial in k[X] factors into a product of linear factors.

**Proof.** See [Dean66].

**B.11.11. Theorem.** Every algebraically closed field k is infinite.

**Proof.** We use Theorem B.8.2. If k has characteristic 0, then the prime subfield of k is isomorphic to the reals. If k has characteristic p, p prime, then one needs to show that the algebraic closure of $\mathbf{Z}_p$ is infinite.

**B.11.12. Theorem.** If k is an algebraically closed field and f is a polynomial in $k[X_1,X_2,\ldots,X_n]$ that vanishes on all but a finite subset of $k^n$, then f is the zero polynomial.

**Proof.** Because k is algebraically closed, it has an infinite number of elements by Theorem B.11.11. If n = 1, then this follows from Theorem B.8.9 since f can only have a finite number of zeros. The general case is proved by induction on n.

**Definition.** If k is a field, then k((X)) will denote the field of quotients of k[[X]]. More generally, $k((X_1,X_2,\ldots,X_n))$ will denote the field of quotients of $k[[X_1,X_2,\ldots,X_n]]$.

There are natural inclusions

$$k \subset k((X_1)) \subset k((X_1,X_2)) \subset \cdots \subset k((X_1,X_2,\ldots,X_n)).$$

The multiplicative inverse of a formal power series f in k[[X]] is 1/f in k((X)). The following fact about 1/f is often used and is therefore worth stating explicitly, namely, 1/f is itself a power series in k[[X]] if it has a nonzero constant term.

**B.11.13. Lemma.** If $f \in k[[X]]$ has a nonzero constant term (equivalently, $\mathrm{ord}(f) = 0$), then there exists a $g \in k[[X]]$ with $fg = 1$.

**Proof.** Simply write out the series expansions for f and g and use the equation $fg = 1$ to solve for the series coefficients of g in terms of those of f.

**B.11.14. Theorem.** The units of $k[[X]]$ are the elements f with $\mathrm{ord}(f) = 0$.

**Proof.** Lemma B.11.13 proved half of the theorem. One still needs to show that every unit has a nonzero constant term. This is left as an easy exercise for the reader.

**B.11.15. Theorem.** Every nonzero element f in $k((X))$ can be written uniquely in the form

$$f = \frac{1}{X^n}(a_0 + a_1X + a_2X^2 + \ldots),$$

where $a_i \in k$ and $a_0 \neq 0$.

**Proof.** See [Walk50].

Theorem B.11.15 leads to a definition of order that extends the one for formal power series.

**Definition.** The integer n in Theorem B.11.15 is called the *order* of f and is denoted by $\mathrm{ord}(f)$.

Theorem B.7.5 can be interpreted as saying that the map $f \to f \circ g$ is a homomorphism of $R[[X]]$ into itself. We can strengthen this in the case where R is field.

**B.11.16. Theorem.** If $f, g \in k[[X]]$, where $\mathrm{ord}(g) = 1$, then the map $f \to f \circ g$ is an order-preserving isomorphism of $k[[X]]$, that is,

(1) $\mathrm{ord}(f) = \mathrm{ord}(f \circ g)$.
(2) There is a $g' \in k[[X]]$ with $\mathrm{ord}(g') = 1$, so that $f = (f \circ g) \circ g'$ for all $f \in k[[X]]$.

**Proof.** See [Walk50].

## B.12 Algebras

**Definition.** A ring A with unity that is a vector space over a field k and that satisfies

$$u(ab) = (ua)b = a(ub), \quad \text{for all} \quad u \in k, a, b \in A,$$

is called an *algebra over k*. If A is also a skew ring, then A is called a *division algebra*.

APPENDIX C

# Basic Linear Algebra

## C.1 More on Linear Independence

Vector spaces were already defined in Appendix B. This appendix gives a highly condensed summary of all the important facts about vector spaces that are used in the book. For more details the interested reader is referred to any book on linear algebra, such as those listed in the bibliography. For simplicity, unless stated otherwise, all vector spaces here are assumed to be finite-dimensional vector spaces over the reals.

Related to the notion of linearly independent vectors is the notion of linearly independent points.

**Definition.** Elements $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_k$ of a vector space are said to be *linearly independent points* if

$$\mathbf{p}_1 - \mathbf{p}_0, \mathbf{p}_2 - \mathbf{p}_0, \ldots, \mathbf{p}_k - \mathbf{p}_0$$

are linearly independent as vectors; otherwise, they are *linearly dependent points*.

**C.1.1. Theorem.** Whether or not points are linearly independent or dependent does not depend on the order in which they are listed.

**Proof.** This is an easy exercise.

In Figure C.1(a), the points $\mathbf{p}_0$, $\mathbf{p}_1$, and $\mathbf{p}_2$ are linearly independent points, but not in Figure C.1(b). Intuitively speaking, points are linearly independent if they generate a maximal dimensional space (maximal with respect to the **number** of points involved). In Figure C.1(a) and (b) the points generate two- and one-dimensional subspaces, respectively. It is because three points **can** generate a two-dimensional space that the points in Figure C.1(b) are called linearly dependent.

Sometimes one wants to decompose a vector space into a sum of subspaces.

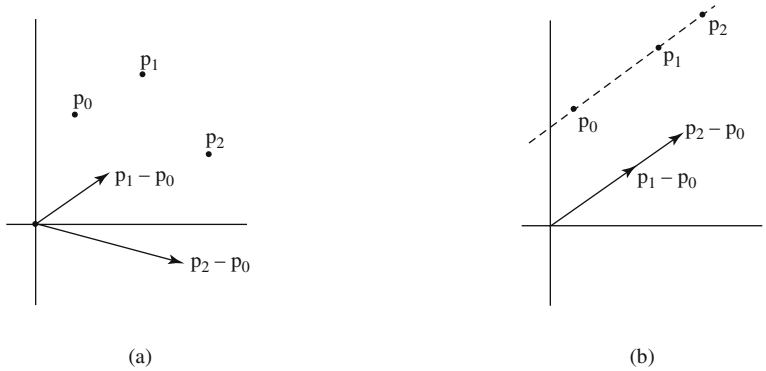



**Figure C.1.** Linearly independent/dependent points.

**Definition.** Let **X** and **Y** be subsets of a vector space **V**. The *sum* **X** + **Y** of **X** and **Y** is defined by

$$\mathbf{X} + \mathbf{Y} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathbf{X} \text{ and } \mathbf{y} \in \mathbf{Y}\}.$$

**C.1.2. Theorem.** If **X** and **Y** are subspaces of a vector space **V**, then **X** + **Y** is a subspace of **V**.

**Proof.** Easy.

**Definition.** A vector space **V** is said to be the *direct sum* of two subspaces **X** and **Y**, and we write

$$\mathbf{V} = \mathbf{X} \oplus \mathbf{Y},$$

if

$$\mathbf{V} = \mathbf{X} + \mathbf{Y} \quad \text{and} \quad \mathbf{X} \cap \mathbf{Y} = \mathbf{0}.$$

**C.1.3. Theorem.** If **X** and **Y** are subspaces of a vector space **V**, then **V** = **X** ⊕ **Y** if and only if each **v** ∈ **V** has a unique representation of the form **v** = **x** + **y**, where **x** ∈ **X** and **y** ∈ **Y**.

**Proof.** See [Lips68].

**Definition.** Let **V** be a vector space and $T: \mathbf{V} \to \mathbf{V}$ a linear transformation. If $T^2 = T$, then T is called a *projection operator* on **V**.

**C.1.4. Theorem.** Let **V** be a vector space. If $T: \mathbf{V} \to \mathbf{V}$ is a projection operator on **V**, then

$$\mathbf{V} = \mathrm{im}(T) \oplus \ker(T).$$

**Proof.** This is an easy consequence of the fact that every $\mathbf{v} \in \mathbf{V}$ can be written in the form

$$\mathbf{v} = T(\mathbf{v}) + (\mathbf{v} - T(\mathbf{v}))$$

and $\mathbf{v} - T(\mathbf{v})$ clearly belongs to $\ker(T)$.

If $\mathbf{V} = \mathbf{X} \oplus \mathbf{Y}$, then the map

$$\begin{aligned} \mathbf{V} &\to \mathbf{X} \\ \mathbf{x} + \mathbf{y} &\to \mathbf{x}, \end{aligned}$$

for $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$, is clearly a projection operator. Therefore, there is a one-to-one correspondence between projection operators on a vector space and direct summands of it.

## C.2 Inner Products

**Definition.** Let $\mathbf{V}_i$, $1 \le i \le n$, and $\mathbf{W}$ be vector spaces over a field k. A map

$$f: \mathbf{V}_1 \times \mathbf{V}_2 \times \cdots \times \mathbf{V}_n \to \mathbf{W}$$

is called a *multilinear map* if, for each i, $1 \le i \le n$,

$$\begin{aligned} f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i + \mathbf{v}_i', \ldots, \mathbf{v}_n) &= f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) + f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_n) \\ f(\mathbf{v}_1, \mathbf{v}_2, \ldots, c\mathbf{v}_i, \ldots, \mathbf{v}_n) &= cf(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n), \end{aligned}$$

for all $\mathbf{v}_j \in \mathbf{V}_j$ and $c \in k$. Equivalently, the map f is multilinear if, for each i and any elements $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_n$ with $\mathbf{v}_j \in \mathbf{V}_j$, the map from $\mathbf{V}_i$ to $\mathbf{W}$ defined by

$$\mathbf{v} \to f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}, \mathbf{v}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_n)$$

is a linear transformation. If $n = 2$, then f is also called a *bilinear map*.

**Definition.** Let $\mathbf{V}$ be a vector space over the field $k = \mathbf{R}$ or $\mathbf{C}$. A bilinear map

$$\begin{aligned} <,>: \ \mathbf{V} \times \mathbf{Y} &\to k \\ (\mathbf{u}, \mathbf{v}) &\to <\mathbf{u}, \mathbf{v}>, \end{aligned}$$

is called an *inner* or *scalar product* on $\mathbf{V}$ if it satisfies the following two additional properties for all $\mathbf{u}, \mathbf{v} \in \mathbf{V}$:

(1) $<\mathbf{u},\mathbf{v}> = \overline{<\mathbf{v},\mathbf{u}>}$ (if k = **R**, then this translates to $<\mathbf{u},\mathbf{v}> = <\mathbf{v},\mathbf{u}>$, that is, <,> is symmetric).
(2) $<\mathbf{u},\mathbf{u}> > 0$ for all nonzero vectors **u**.

An *inner product space* is a pair (**V**,<,>), where **V** is a vector space and <,> is an inner product on **V**.

Note that $<\mathbf{u},\mathbf{u}>$ is always a real number by condition (1), so that (2) makes sense. It is easy to show that a general inner product also satisfies

$$<\mathbf{0},\mathbf{u}> = 0, \text{ for all } \mathbf{u},$$

and

$$<\mathbf{u},\mathbf{u}> = 0 \text{ if and only if } \mathbf{u} = 0.$$

**Definition.** The (standard) *dot product on* $\boldsymbol{R}^n$ is defined by

$$\mathbf{u} \bullet \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

**Definition.** The (standard) *dot product on* $\boldsymbol{C}^n$ is defined by

$$\mathbf{u} \bullet \mathbf{v} = u_1\overline{v_1} + u_2\overline{v_2} + \cdots + u_n\overline{v_n}.$$

**C.2.1. Theorem.** The dot products on $\mathbf{R}^n$ and $\mathbf{C}^n$ are inner products.

**Proof.** It is easy to check that the properties in the definition of a dot product are satisfied.

If all one wants to do is to have a dot product in $\mathbf{R}^n$, we could have dispensed with the definition of an inner product and simply shown that the dot product satisfies the properties listed in the definition. However, the point to abstracting the basic properties into a definition is that it isolates the essential properties of an inner product and one does not get sidetracked by details. Vector spaces admit many **different** functions that satisfy the definition of an inner product.

An inner product on a vector space enables us to give a simple definition of the length of a vector.

**Definition.** Let **V** be a vector space with an inner product <,> and let $\mathbf{v} \in \mathbf{V}$. Define the *length* $|\mathbf{v}|$ of **v** by

$$|\mathbf{v}| = \sqrt{<\mathbf{v},\mathbf{v}>}.$$

A vector of length 1 is called a *unit* vector.

If one writes out this definition of the length of a vector for the standard dot product on $\mathbf{R}^n$ in terms of its coordinates, one sees that it is just the usual Euclidean length; however, that is not the really important point. It is property (2) in the definition of a dot product that guarantees that our definition of length is well defined. It

and the other properties also guarantee that the triangle inequality is satisfied (see Theorem C.2.2 below), so that we have a **good** definition of length. This aspect needs to be emphasized. What makes linear algebra beautiful is that it enables us to solve problems in an elegant, clean way without having to get involved in messy computations with coordinates. As long as we use only general (and **essential**) properties like those in the definition for an inner product, we shall be able to give simple proofs.

**Definition.** Let **V** be a vector space. Given two points $\mathbf{p},\mathbf{q} \in \mathbf{V}$, define the *vector from* ***p*** *to* ***q***, **pq**, by

$$\mathbf{pq} = \mathbf{q} - \mathbf{p}.$$

If **V** has an inner product, then define the *distance from* ***p*** *to* ***q***, dist (**p**,**q**), by

$$\text{dist}(\mathbf{p},\mathbf{q}) = |\mathbf{pq}|.$$

There are two very important inequalities.

**C.2.2. Theorem.** Let **u** and **v** be vectors in a vector space with an inner product <,>.

(1) (The Cauchy-Schwarz inequality) $|<\mathbf{u},\mathbf{v}>| \le |\mathbf{u}|\ |\mathbf{v}|$

We have equality if and only if **u** and **v** are linearly dependent.

(2) (The triangle inequality) $|\mathbf{u} + \mathbf{v}| \le |\mathbf{u}| + |\mathbf{v}|$

We have equality **only if u** and **v** are linearly dependent.

**Proof.** We prove (1) first. Let c be any scalar. Then

$$0 \le <\mathbf{u} - c\mathbf{v}, \mathbf{u} - c\mathbf{v}> = |\mathbf{u}|^2 - 2c <\mathbf{u}, \mathbf{v}> + c^2|\mathbf{v}|^2. \tag{C.1}$$

If we consider the right-hand side of (C.1) as a quadratic equation in the variable c, then we can use the discriminant test from the quadratic formula to conclude that "<" holds (that is, there are no solutions) if and only if

$$[-2 <\mathbf{u}, \mathbf{v}>]^2 - 4|\mathbf{u}|^2|\mathbf{v}|^2 < 0,$$

which simplifies to what we want. On the other hand, it is easy to see from (C.1) that equality holds if and only if $\mathbf{u} = c\mathbf{v}$ or $\mathbf{v} = \mathbf{0}$ (that is, **u** and **v** are linearly dependent).

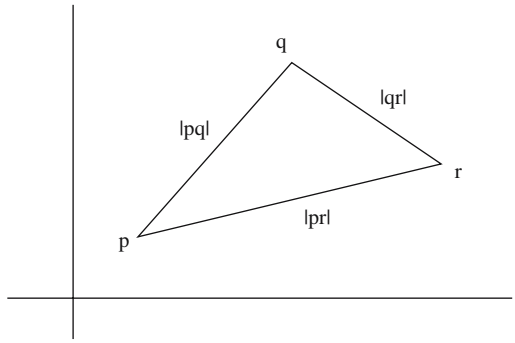
An alternate way to prove the Cauchy-Schwarz inequality is simply to set c to

$$\frac{1}{|\mathbf{v}|^2} <\mathbf{u}, \mathbf{v}>$$

in (C.1) and simplify the resulting expression.

Part (2) of the theorem follows from the Cauchy-Schwarz inequality because

**Figure C.2.** The triangle inequality.




$$
\begin{aligned}
|\mathbf{u}+\mathbf{v}|^2 &= <\mathbf{u}+\mathbf{v},\mathbf{u}+\mathbf{v}> \\
&= |\mathbf{u}|^2 + 2<\mathbf{u},\mathbf{v}> + |\mathbf{v}|^2 \\
&\le |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}| + |\mathbf{v}|^2 \\
&= (|\mathbf{u}|+|\mathbf{v}|)^2
\end{aligned}
$$

and equality holds only if **u** and **v** are linearly dependent.

The geometric content of the triangle inequality is that the sum of the lengths of two sides of a triangle is larger than the length of the third side (see Figure C.2) and is summarized in the next corollary.

**C.2.3. Corollary.** If $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbf{R}^n$, then

$$|\mathbf{pr}| < |\mathbf{pq}| + |\mathbf{qr}|$$

unless **p** and **q** and **r** are collinear.

## C.3 Matrices of Linear Transformations

We begin with a brief summary of basic facts dealing with matrices. See, for example, [John67], [Lips68], or [NobD77] for more details of proofs.

**Definition.** An $n \times n$ matrix $(a_{ij})$ over the reals is said to be *symmetric* if $a_{ji} = a_{ij}$ for all i and j. An $n \times n$ matrix $(a_{ij})$ over the complex numbers is said to be *Hermitian* if

$$a_{ji} = \overline{a_{ij}}$$

for all i and j. An arbitrary $n \times n$ matrix $(a_{ij})$ is said to be *diagonal* if $a_{ij} = 0$ for all $i \neq j$.

**Definition.** Let $a = (a_{ij})$ be an $n \times n$ matrix. The *determinant* of A, denoted by det(A) or $|A|$, is defined by

$$\det(A) = \sum_{\sigma \in S_n} (\operatorname{sign}(\sigma))\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Let $M_{ij}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column. The determinant of $M_{ij}$ is called a *minor* of A. The *ijth cofactor* of A, $A_{ij}$, is the signed minor defined by

$$A_{ij} = (-1)^{i+j} |M_{ij}|.$$

The $n \times n$ matrix $(A_{ij})$ of cofactors is called the *adjoint matrix* of A and is denoted by adj(A).

**C.3.1. Theorem.** The determinant function satisfies the following properties:

(1) $\det A^T = \det A$.
(2) If the matrix B is obtained from the matrix A by interchanging two rows, then $\det B = -\det A$.
(3) If a matrix A has two identical rows, then $\det A = 0$.
(4) If a matrix A has a row of zeros, then $\det A = 0$.
(5) Assume that the matrices A, A′, and A″ are identical except for the ith rows $A_i$, $A'_i$, and $A''_i$, respectively. Assume further that $A_i = aA'_i + bA''_i$. Then

$$\det A = a \det A' + b \det A''.$$

(6) $\det AB = (\det A)(\det B)$.
(7) $\det A^{-1} = 1/(\det A)$.
(8) The determinant of a matrix is sometimes usefully computed by means of "expansion by minors," that is, if $A = (a_{ij})$ and if $A_{ij}$ is the ijth cofactor of A, then

$$|A| = \sum_{j=1}^{n} a_{ij} A_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}.$$

(9) The *inverse* $A^{-1}$ of a matrix A is defined by the equations $AA^{-1} = A^{-1}A = I$. It can be computed by means of the determinant and the adjoint matrix, that is,

$$A^{-1} = \frac{1}{|A|} \operatorname{adj} A.$$

(10) A matrix has an inverse, or is *nonsingular*, if and only if it has a nonzero determinant. (A matrix without an inverse is said to be *singular*.)

**Proof.** See [Lips68].

**Definition.** If $A = (a_{ij})$ is an $n \times n$ matrix, then the *trace* of A, denoted by tr(A), is defined by

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

**C.3.2. Theorem.** The trace function satisfies the following properties:

(1) $\mathrm{tr}(aA + bB) = a\ \mathrm{tr}\ A + b\ \mathrm{tr}\ B$.
(2) $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.
(3) $\mathrm{tr}(A) = \mathrm{tr}(P^{-1}AP)$ for any nonsingular matrix P.

**Proof.** Parts (1) and (2) follow from some simple computations. Part (3) follows from (2).

**Definition.** Let $A = (a_{ij})$ be an $n \times m$ matrix. The rows of A can be thought of as vectors in $\mathbf{R}^m$. The *row rank* of A is the dimension of the subspace in $\mathbf{R}^m$ that these vectors generate. Similarly, the columns of A can be thought of as vectors in $\mathbf{R}^n$. The *column rank* of A is the dimension of the subspace in $\mathbf{R}^n$ that these vectors generate.

One can show that the row rank and column rank of a matrix are the same.

**Definition.** The *rank* of a matrix is the common value of the row rank or column rank. An $n \times m$ matrix has *maximal rank* if its rank is the smaller of n or m.

**C.3.3. Theorem.** The rank of a matrix is the dimension of its largest nonsingular square submatrix. An $n \times n$ matrix is nonsingular if and only if it has rank n.

**Proof.** See [John67].

It is assumed that the reader is familiar how matrices are used to solve linear systems of equations of the form

$$A\mathbf{x}^T = \mathbf{b}^T,$$

in particular the method of Gauss elimination. (We need the transpose of the vectors because in this book vectors in $\mathbf{R}^n$ are treated as $1 \times n$ matrices.) We will not describe the method here, but there is some terminology that one runs into when the method is discussed, which we want to record for the sake of completeness. Recall that if Gauss elimination, applied to a matrix A to get an upper triangular matrix U, does not involve interchanging rows, then A can be written in the form

$$A = LU,$$

where L is a lower-triangular matrix and U is an upper-triangular matrix. This reduces the system of equations above to the two systems

$$L\mathbf{y}^T = \mathbf{b}^T$$

and

$$U\mathbf{x}^T = \mathbf{y}^T,$$

which are easy to solve. If row interchanges are involved, then another factor comes into play and A can be written in the form

$$A = PLU,$$

where P is a *permutation matrix*, that is, a matrix obtained from an identity matrix by a sequence of column and/or row interchanges. See [NobD77].

**Definition.** The representations A = LU or A = PLU are called *LU-decompositions* of A.

Next, let $T: \mathbf{V} \to \mathbf{V}$ be a linear transformation and assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for $\mathbf{V}$. If $\mathbf{v} \in \mathbf{V}$, then

$$T(\mathbf{v}_i) = \sum_{j=1}^{n} a_{ij}\mathbf{v}_j.$$

The fact that the vectors $\mathbf{v}_i$ form a basis implies that the coefficients $a_{ij}$ are unique.

**Definition.** The matrix $a = (a_{ij})$ is called the *matrix for the linear transformation* T with respect to the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

Clearly, the matrix for a linear transformation depends on the basis of the vector space that is used in the definition. It is easy to describe this dependence.

**Definition.** Two $n \times n$ matrices A and B are *similar* if there exists a nonsingular matrix P so that

$$B = P^{-1}AP.$$

Similarity of matrices is an equivalence relation.

**C.3.4. Theorem.** Let A be the matrix for a linear transformation T with respect to a given basis. A matrix B represents T with respect to some other basis if and only if B is similar to A.

**Proof.** See [John67].

One reason for defining a matrix for a linear transformation is that it allows us to evaluate that transformation using matrix multiplication. It is very important that one use the correct matrix and not its transpose. With our choice and the fact that our vectors in $\mathbf{R}^n$ are **row** vectors ($1 \times n$ matrices), given a linear transformation

$$T: \mathbf{R}^n \to \mathbf{R}^n, \tag{C.2a}$$

then

$$T(\mathbf{p}) = \mathbf{p}A. \tag{C.2b}$$

If we had chosen the transpose of A, then

$$T(\mathbf{p}) = \left(A\mathbf{p}^T\right)^T.$$

The distinction between choosing A or its transpose is therefore a question of whether we want to pre- or post-multiply vectors by matrices. It does not matter which one chooses as long as one is **consistent**. In order to have the matrix product agree with the action of the map, the reader needs to take note of the following:

**Our choice of matrices is such that one must always pre-multiply vectors!**

In this way we avoid excessive transpose operations in the writing of formulas. (They would be needed because there is a difference between a row and a column vector. Our vectors are row vectors.)

**Important Note!** Unless stated otherwise, the matrix for a linear transformation $T:\mathbf{R}^n \to \mathbf{R}^n$ will always be defined **with respect to the standard basis**. Furthermore, with this assumption one can then use equation (C.2b) to define an **unambiguous bijective correspondence** between matrices and such linear transformations. This is the correspondence we will have in mind if we have the need to pass back and forth between matrices and transformations. A similar comment applies to transformations $T:k^n \to k^n$ for some field k.

**C.3.5. Theorem.** Let T, $T_1$, $T_2:\mathbf{V} \to \mathbf{V}$ be linear transformations and assume that A, $A_1$, and $A_2$ are the matrices for T, $T_1$, and $T_2$, respectively, with respect to some fixed basis for **V**.

(1) The matrix for $T^{-1}$ is $A^{-1}$.
(2) if $S = T_1T_2$, then $A_2A_1$ is the matrix for S.

**Proof.** The proofs follow by straightforward computations. Note though that because of our conventions the matrices in (2) are listed in the opposite order from that of the transformations!

**Definition.** If A is the matrix for the linear transformation T, then the determinant of A is called the *determinant* of T and is denoted by det(T). The trace of A is called the *trace* of T and is denoted by tr(T). The rank of A is called the *rank* of T.

**C.3.6. Theorem.** The determinant, trace, and rank of a linear transformation $T:\mathbf{V} \to \mathbf{V}$ depends only on T and not the choice of basis for **V**.

**Proof.** The theorem is an easy consequence of Theorem C.3.4, property (6) of determinants in Theorem C.3.1, and property (2) of the trace function in Theorem C.3.2.

**C.3.7. Theorem.** A linear transformation $T:\mathbf{V} \to \mathbf{V}$ is nonsingular if and only if $\det(T) \neq 0$.

**Proof.** See [John67].

## C.4 Eigenvalues and Eigenvectors

Let $\mathbf{V}$ be a vector space over a field k and $T:\mathbf{V} \to \mathbf{V}$ a linear transformation.

**Definition.** A scalar $\lambda$ in k is called an *eigenvalue* of T if there exists a nonzero vector $\mathbf{v}$ in $\mathbf{V}$ such that

$$T(\mathbf{v}) = \lambda\mathbf{v}. \tag{C.3}$$

Every vector satisfying equation (C.3) is called an *eigenvector* for the eigenvalue $\lambda$. The set of eigenvectors for an eigenvalue $\lambda$ is called the *eigenspace* of $\lambda$. If A is an $n \times n$ matrix over k, then the *eigenvalues* and *eigenvectors of A* are the eigenvalues and eigenvectors, respectively, of the linear transformation $T:k^n \to k^n$ associated to A.

**C.4.1. Lemma.** The eigenspace of an eigenvalue $\lambda$ is the kernel of the transformation $T - \lambda I$ and hence is a vector subspace.

**Proof.** Straightforward.

**Definition.** A linear transformation $T:\mathbf{V} \to \mathbf{V}$ that can be represented by a diagonal matrix with respect to some basis of $\mathbf{V}$ is said to be *diagonalized* by the basis, or simply *diagonalizable*. An $n \times n$ matrix is *diagonalizable* if the associated linear transformation on $k^n$ is diagonalizable.

**C.4.2. Theorem.** A linear transformation $T:\mathbf{V} \to \mathbf{V}$ is diagonalizable if and only if $\mathbf{V}$ has a basis of eigenvectors of T. The diagonal entries of the matrix with respect to the basis that diagonalizes the transformation are then the eigenvalues of T.

**Proof.** Easy.

**C.4.3. Theorem.** Let $T:\mathbf{V} \to \mathbf{V}$ be a linear transformation. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are nonzero eigenvectors for T corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, respectively, then the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly independent.

**Proof.** The proof is by induction on m. The case m = 1 is clear. Assume that the theorem has been proved for m − 1, m > 1. Assume that the m eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ satisfy a relation

$$\mathbf{0} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m, \tag{C.4}$$

for some $a_i$. Applying T to both sides gives

$$\begin{aligned}\mathbf{0} &= T(\mathbf{0})\\ &= a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) + \cdots + a_mT(\mathbf{v}_m)\\ &= a_1\lambda_1v_1 + a_2\lambda_2v_2 + \cdots + a_m\lambda_mv_m\end{aligned} \tag{C.5}$$

On the other hand, multiplying (C.4) by $\lambda_m$ and subtracting from (C.5) gives

$$\mathbf{0} = a_1(\lambda_1 - \lambda_m)\mathbf{v}_1 + a_2(\lambda_2 - \lambda_m)\mathbf{v}_2 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)\mathbf{v}_{m-1}.$$

By our inductive hypothesis we must have

$$a_1 = a_2 = \cdots = a_{m-1} = 0,$$

since none of the terms $\lambda_i - \lambda_m$, $i = 1, 2, \ldots m - 1$, are zero. This and (C.4) in turn implies that $a_m = 0$, and we are done.

**C.4.4. Corollary.** Let $\mathbf{V}$ be an n-dimensional vector space. If a linear transformation $T:\mathbf{V} \to \mathbf{V}$ has n distinct eigenvalues, then it is diagonalizable.

Because of Theorem C.4.2 and the fact that diagonalizable linear transformations are easy to understand, finding the eigenvalues and eigenvectors of a linear transformation is an important problem.

**C.4.5. Theorem.** A scalar $\lambda$ is an eigenvalue of a linear transformation $T:\mathbf{V} \to \mathbf{V}$ if and only if the transformation $T - \lambda I$ is singular.

**Proof.** Straightforward.

**Definition.** Let A be an $n \times n$ matrix. The polynomial

$$\det(tI_n - A), \tag{C.6}$$

where $I_n$ is the $n \times n$ identity matrix, is called the *characteristic polynomial of A*. If $T:\mathbf{V} \to \mathbf{V}$ is a linear transformation and A is a matrix that represents T with respect to some basis of $\mathbf{V}$, then the characteristic polynomial of A is called the *characteristic polynomial of T*.

**C.4.6. Lemma.** The characteristic polynomial of a transformation is well defined and independent of the matrix that is chosen to represent it.

**Proof.** Use Theorem C.3.4 and properties of the determinant.

**C.4.7. Theorem.** Let $T:\mathbf{V} \to \mathbf{V}$ be a linear transformation. A scalar $\lambda$ is an eigenvalue of T if and only if $\lambda$ is a root of the characteristic polynomial of T.

**Proof.** Use Theorem C.4.5 and Theorem C.3.1(10).

**C.4.8. Example.** To analyze the linear transformation $T:\mathbf{R}^2 \to \mathbf{R}^2$ defined by the matrix

$$A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix}$$

in terms of eigenvalues, eigenvectors, and eigenspaces.

**Solution.** We need to solve the equation

$$|\lambda I - A| = \lambda^2 - 4 = 0$$

for $\lambda$. The solutions are $\lambda = 2$ or $-2$, which are the eigenvalues of T. Next, to find the eigenvectors, we need to solve

$$(x\ y)A = \lambda(x\ y),$$

that is,

$$\begin{aligned} (1-\lambda)x + \qquad y &= 0 \\ 3x - (1+\lambda)y &= 0. \end{aligned}$$

When $\lambda = 2$, this leads to $x = y$. In other words, (1,1) is an eigenvector for the eigenvalue 2 and is a basis for its eigenspace. Similarly, when $\lambda = -2$, then $y = -3x$, and (1,–3) is an eigenvector for the eigenvalue –2 and is a basis for its eigenspace.

We can use Theorem C.4.7 to show that not every linear transformation on a vector space over the reals is diagonalizable. For example, consider the transformation on $\mathbf{R}^2$ represented by the matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The characteristic polynomial of this transformation is $t^2 + 1$, which has no real root. Of course, over the complex numbers every polynomial has a root, so that linear transformations over complex vector spaces always have eigenvalues and eigenvectors.

The next proposition lists two properties of the characteristic polynomial which sometimes come in handy, especially in the two-dimensional case where they completely characterize the polynomial.

**C.4.9. Proposition.** Let $A = (a_{ij})$ be an $n \times n$ matrix and let

$$p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$$

be its characteristic polynomial. Then

(1) $a_{n-1} = -\mathrm{tr}\,(A)$,
(2) $a_0 = (-1)^n \det\,(A)$, and
(3) if $n = 2$, then

$$p(t) = t^2 - (\mathrm{tr}(A))t + \det(A).$$

**Proof.** Using properties of the determinant it is easy to see that

$$p(t) = (t - a_{11})(t - a_{22}) \cdots (t - a_{nn}) + \text{polynomial in } t \text{ of degree } n - 2.$$

This equation and the definition of the trace function implies (1). To prove (2), simply substitute 0 for t in (C.6). Part (3) follows immediately from (1) and (2).

Here is a fundamental theorem about the diagonalizability of transformations.

**C.4.10. Theorem.** Let $T:\mathbf{V} \to \mathbf{V}$ be a linear transformation. Assume that $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the distinct eigenvalues of T and let $d_1, d_2, \ldots, d_m$ be the dimensions of the corresponding eigenspaces. Let p(t) be the characteristic polynomial of T. Then T is diagonalizable if and only if

$$p(t) = (t-\lambda_1)^{d_1}(t-\lambda_2)^{d_2}\cdots(t-\lambda_m)^{d_m}.$$

**Proof.** This follows easily from Theorem C.4.2 and Theorem C.4.7.

Much more can be said about the diagonalizability of transformations. For example, see [HofK71].

## C.5 The Dual Space

Given vector spaces **V** and **W** over a field k, let

$$L(\mathbf{V},\mathbf{W}) = \{T : \mathbf{V} \to \mathbf{W} \mid T \text{ is a linear transformation}\}.$$

If S, T $\in$ L(**V**,**W**) and a $\in$ k, then define

$$S+T,\ aS:\ \mathbf{V} \to \mathbf{W}$$

by

$$(S+T)(\mathbf{v}) = S(\mathbf{v}) + T(\mathbf{v})$$
$$(aS)(\mathbf{v}) = a(S(\mathbf{v})).$$

**C.5.1. Theorem.** The maps S + T and aS are linear transformation and this addition and scalar multiplication make L(**V**,**W**) into a vector space over k.

**Proof.** Easy.

**Definition.** Let **V** be a vector space over a field k. A linear transformation $T:\mathbf{V} \to k$ is called a *linear functional* on **V**. The vector space of linear functionals on **V** is called the *dual space* of **V** and is denoted by **V***.

Let **V** be a vector space over a field k and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis for **V**. Define linear functionals

$$\mathbf{v}_i^* :\ \mathbf{V} \to \mathbf{R}$$

by

$$\mathbf{v}_i^*(\mathbf{v}_j) = \delta_{ij}.$$

**C.5.2. Theorem.** The map from $\mathbf{V}$ to $\mathbf{V}^*$ that sends $\mathbf{v}_i$ to $\mathbf{v}_i^*$ is a vector space isomorphism.

**Proof.** Straightforward.

The isomorphism in Theorem C.5.2 between $\mathbf{V}$ and $\mathbf{V}^*$ clearly depends on the basis.

**Definition.** The basis $\mathbf{v}_1^*, \mathbf{v}_2^*, \ldots, \mathbf{v}_n^*$ is called the *dual basis* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

Notice that if

$$\mathbf{w} = \sum_{i=1}^{n} a_i \mathbf{v}_i,$$

then $\mathbf{v}_i^*(\mathbf{w}) = a_i$, so that the ith dual basis element just picks out the ith component coefficient of the expansion of a vector $\mathbf{w}$ in terms of the $\mathbf{v}_i$'s.

**C.5.3. Theorem.** Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces over a field k and let $T: \mathbf{V} \to \mathbf{W}$ be a linear transformation. The map

$$T^*: \mathbf{W}^* \to \mathbf{V}^*$$

defined by

$$T^*(\alpha)(\mathbf{v}) = \alpha(T(\mathbf{v})), \text{ for } \mathbf{v} \in \mathbf{V},$$

is a linear transformation.

**Proof.** Easy.

**Definition.** The map $T^*$ in Theorem C.5.3 is called the *dual map* of the linear transformation T.

Next, given $\mathbf{v} \in \mathbf{V}$, define $\mathbf{v}^{**} \in \mathbf{V}^{**} = (\mathbf{V}^*)^*$ by

$$\mathbf{v}^{**}(\alpha) = \alpha(\mathbf{v}), \text{ for } \alpha \in \mathbf{V}^*.$$

**C.5.4. Theorem.** The map from $\mathbf{V}$ to $\mathbf{V}^{**}$ that sends $\mathbf{v}$ to $\mathbf{v}^{**}$ is a vector space isomorphism called the *natural isomorphism* between $\mathbf{V}$ and $\mathbf{V}^{**}$.

**Proof.** Easy.

Note that, although the isomorphism between $\mathbf{V}$ and $\mathbf{V}^*$ depended on the choice of a basis for $\mathbf{V}$, the isomorphism between $\mathbf{V}$ and $\mathbf{V}^{**}$ does not. This allows us to identify $\mathbf{V}^{**}$ with $\mathbf{V}$ in a natural way and one often makes this identification.

# C.6 The Tensor and Exterior Algebra

This section contains some rather technical mathematics. It is needed because without a notion of tensor and exterior algebras one cannot discuss differential forms in a rigorous way. We could have simplified some of the discussion by restricting ourselves to the algebras of multilinear and alternating multilinear maps as is done, for example, by Spivak in [Spiv65] and [Spiv70a]. Those algebras will be highlighted in the discussion below, but they are special cases of a general construction and by considering only them the reader would have gotten an incomplete picture of the subject. For that reason, we considered it worthwhile to outline the "correct" development of these algebras. The reader still has the choice of skipping uninteresting material. Other references for tensor and exterior algebras are [AusM63] and [KobN63].

In this section the field for a vector space will always be the reals $\mathbf{R}$.

**Notation.** Let $\mathbf{V}_i$, $1 \le i \le k$, and $\mathbf{W}$ be vector spaces. The set of multilinear maps

$$f: \mathbf{V}_1 \times \mathbf{V}_2 \times \cdots \times \mathbf{V}_k \to \mathbf{W}$$

will be denoted by $L^k(\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_k;\mathbf{W})$. If $\mathbf{V} = \mathbf{V}_1 = \mathbf{V}_2 = \ldots = \mathbf{V}_k$, then $L^k(\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_k;\mathbf{W})$ will be abbreviated to $L^k(\mathbf{V};\mathbf{W})$ and $L^k(\mathbf{V};\mathbf{R})$ is abbreviated to $L^k(\mathbf{V})$. It follows that $L^k(\mathbf{V})$ denotes the multilinear maps

$$f: \mathbf{V}^k = \underbrace{\mathbf{V} \times \mathbf{V} \times \cdots \times \mathbf{V}}_{k} \to \mathbf{R}.$$

By setting $\mathbf{V}^0 = \mathbf{R}$, the notation $L^0(\mathbf{V})$ makes sense and denotes the set of linear maps $\mathbf{R} \to \mathbf{R}$.

We may always identify $L^0(\mathbf{V})$ with $\mathbf{R}$ using the natural map $L^0(\mathbf{V}) \to \mathbf{R}$ that sends $\alpha$ to $\alpha(1)$.

**Definition.** A *tensor product* of vector spaces $\mathbf{V}_i$, $1 \le i \le k$, is a pair $(\mathbf{A},\alpha)$, where $\mathbf{A}$ is a vector space and

$$\alpha: \mathbf{V}_1 \times \mathbf{V}_2 \times \cdots \times \mathbf{V}_k \to \mathbf{A}$$

is a multilinear map that satisfy the following *universal factorization property*:

If $f: \mathbf{V}_1 \times \mathbf{V}_2 \times \ldots \times \mathbf{V}_k \to \mathbf{W}$ is any multilinear map into a vector space $\mathbf{W}$, then there is a **unique** linear transformation $g: \mathbf{A} \to \mathbf{W}$ so that $f = g \circ \alpha$. See Figure C.3.

$$\begin{tikzcd}
V_1 \times V_2 \times \cdots \times V_k \arrow[r, "\alpha"] \arrow[dr, "f"'] & A \arrow[d, dashed, "g"] \\
& W
\end{tikzcd}$$

**Figure C.3.** The universal factorization property of tensor products.

**C.6.1. Theorem.** Let $\mathbf{V}_i$, $1 \le i \le k$, be vector spaces.

(1) (Existence) A tensor product $(\mathbf{A},\alpha)$ of the $\mathbf{V}_i$ exists. The image of $\alpha$ will actually span $\mathbf{A}$.
(2) (Uniqueness) Given another tensor product $(\mathbf{B},\beta)$, then there is a unique isomorphism $\mu:\mathbf{A} \to \mathbf{B}$ with $\beta = \mu \circ \alpha$.

**Proof.** To prove part (1), here is how one can define $\mathbf{A}$ and $\alpha$ when $k = 2$. It should be obvious how to generalize the construction to handle the case of an arbitrary k. Let $\mathbf{M}$ be the free vector space with basis $(\mathbf{v}_1,\mathbf{v}_2)$, where $\mathbf{v}_i \in \mathbf{V}_i$. Let $\mathbf{N}$ be the vector subspace of $\mathbf{M}$ generated by all elements of $\mathbf{M}$ of the form

$$(\mathbf{v}_1 + \mathbf{v}_1', \mathbf{v}_2) - (\mathbf{v}_1, \mathbf{v}_2) - (\mathbf{v}_1', \mathbf{v}_2),\ (\mathbf{v}_1, \mathbf{v}_2 + \mathbf{v}_2') - (\mathbf{v}_1, \mathbf{v}_2) - (\mathbf{v}_1, \mathbf{v}_2'),$$
$$(r\mathbf{v}_1, \mathbf{v}_2) - r(\mathbf{v}_1, \mathbf{v}_2),\ (\mathbf{v}_1, r\mathbf{v}_2) - r(\mathbf{v}_1, \mathbf{v}_2),$$

where $\mathbf{v}_i$, $\mathbf{v}_i' \in \mathbf{V}_i$ and $r \in \mathbf{R}$. Define $A = \mathbf{M}/\mathbf{N}$ and

$$\alpha:\ \mathbf{V}_1 \times \mathbf{V}_2 \to \mathbf{A}$$

by

$$\alpha(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1, \mathbf{v}_2) + \mathbf{N}.$$

Given a map $f:\mathbf{V}_1 \times \mathbf{V}_2 \to \mathbf{W}$, define $g_0:\mathbf{M} \to \mathbf{W}$ by

$$g_0\left(\sum_{(\mathbf{v}_1,\mathbf{v}_2)\in \mathbf{V}_1\times\mathbf{V}_2} a_{\mathbf{v}_1,\mathbf{v}_2}(\mathbf{v}_1,\mathbf{v}_2)\right) = \sum_{(\mathbf{v}_1,\mathbf{v}_2)\in \mathbf{V}_1\times\mathbf{V}_2} a_{\mathbf{v}_1,\mathbf{v}_2} f(\mathbf{v}_1,\mathbf{v}_2).$$

One can show that $g_0$ sends $\mathbf{N}$ to $\mathbf{0}$ and hence induces a map $g:\mathbf{A} \to \mathbf{W}$. Clearly, $f = g \circ \alpha$. The map g is unique because the image of $\alpha$ spans $\mathbf{A}$. It follows that the pair $(\mathbf{A},\alpha)$ is a tensor product for $\mathbf{V}_1$ and $\mathbf{V}_2$. See [AusM63].

To prove the uniqueness of the tensor product, let $(\mathbf{B},\beta)$ be another such. See Figure C.4. Since $\beta$ is multilinear, the universal factorization property of $(\mathbf{A},\alpha)$ implies that there is a unique linear map $\mu:\mathbf{A} \to \mathbf{B}$ with $\beta = \mu \circ \alpha$. Similarly, the universal factorization property of $(\mathbf{B},\beta)$ implies that there is a unique linear map $\mu':\mathbf{B} \to \mathbf{A}$ with $\alpha = \mu' \circ \beta$. Therefore, $\beta = \mu \circ \mu' \circ \beta$ and $\alpha = \mu' \circ \mu \circ \alpha$. This implies that $\mu$ is an isomorphism. The theorem is proved.

Theorem C.6.1(2) shows that it is the universal factorization property of a tensor product that is important and not the particular construction that is used. For that reason one usually talks about "the" tensor product and uses a uniform notation.

$$\begin{tikzcd}
V_1 \times V_2 \times \cdots \times V_k \arrow[rr, "\alpha"] \arrow[dr, "\beta"'] & & A \arrow[dl, shift left, "\mu"] \\
& B \arrow[ur, shift left, "\mu'"] &
\end{tikzcd}$$

**Figure C.4.** The uniqueness of tensor products.

**Notation.** Let $(\mathbf{A},\alpha)$ be the tensor product of vector spaces $\mathbf{V}_i$, $1 \le i \le k$, constructed as in Theorem C.6.1. The space $\mathbf{A}$ will be denoted by $\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \ldots \otimes \mathbf{V}_k$ and the map $\alpha$ by $\otimes$. If $\mathbf{v}_i \in \mathbf{V}_i$, then the element $\otimes(\mathbf{v}_1,\mathbf{v}_2, \ldots ,\mathbf{v}_k)$ in $\mathbf{A}$ will be denoted by $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_k$ and is called the *tensor product* of the vectors $\mathbf{v}_i$.

We summarize some basic properties of the tensor product. Part of what we have accomplished is that we have formalized a tensor product notation for vectors.

**C.6.2. Theorem.** Let $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ be vector spaces.

(1) Let $\mathbf{u}$, $\mathbf{u}_i \in \mathbf{U}$, $\mathbf{v}$, $\mathbf{v}_i \in \mathbf{V}$, and $c_i \in \mathbf{R}$. Then

$$(c_1\mathbf{u}_1 + c_2\mathbf{u}_2) \otimes \mathbf{v} = c_1(\mathbf{u}_1 \otimes \mathbf{v}) + c_2(\mathbf{u}_2 \otimes \mathbf{v})$$
$$\mathbf{u} \otimes (c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1(\mathbf{u} \otimes \mathbf{v}_1) + c_2(\mathbf{u} \otimes \mathbf{v}_2)$$

(2) The map

$$c \otimes \mathbf{v} \to c\mathbf{v}$$

induces a natural isomorphism

$$R \otimes \mathbf{V} \to \mathbf{V}.$$

Using this isomorphism, we shall always identify $\mathbf{R} \otimes \mathbf{V}$ with $\mathbf{V}$.

(3) (Associativity) The maps

$$(\mathbf{u} \otimes \mathbf{v}) \otimes \mathbf{w} \to \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$$
$$\mathbf{u} \otimes (\mathbf{v} \otimes \mathbf{w}) \to \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$$

induce natural isomorphisms

$$(\mathbf{U} \otimes \mathbf{V}) \otimes \mathbf{W} \to \mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}$$
$$\mathbf{U} \otimes (\mathbf{V} \otimes \mathbf{W}) \to \mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W},$$

respectively. As a result one does not have to worry about parenthesizing tensor products.

(4) If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are bases for $\mathbf{U}$ and $\mathbf{V}$, respectively, then the $\mathbf{u}_i \otimes \mathbf{v}_j$ for $1 \le i \le n$ and $1 \le j \le m$ form a basis for $\mathbf{U} \otimes \mathbf{V}$. In particular,

$$\dim(\mathbf{U} \otimes \mathbf{V}) = (\dim \mathbf{U})(\dim \mathbf{V}).$$

**Proof.** The proofs are straightforward and easy. See [AusM63].

Note another property of the tensor product: there is a one-to-one correspondence between multilinear maps from $\mathbf{V}_1 \times \mathbf{V}_2 \times \ldots \times \mathbf{V}_k$ to a vector space $\mathbf{W}$ and **linear** maps from $\mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \ldots \otimes \mathbf{V}_k$ to $\mathbf{W}$. In other words, rather than talking about multilinear

maps we can always talk about ordinary linear maps instead. We shall see this in action below but introduce some notation first.

**Definition.** Let $\mathbf{V}$ be a vector space. Define the *r-fold tensor product* of $\mathbf{V}$, denoted by $T^k\mathbf{V}$, by

$$T^0\mathbf{V} = \mathbf{R},$$
$$T^k\mathbf{V} = \underbrace{\mathbf{V} \otimes \mathbf{V} \otimes \cdots \otimes \mathbf{V}}_{k}, \quad k \geq 1.$$

Let $T\mathbf{V}$ denote the direct sum of the vector spaces $T^k\mathbf{V}$, $k \geq 0$. The product operation

$$\otimes : \; T^r\,\mathbf{V} \times T^s\mathbf{V} \to T^{r+s}\mathbf{V}$$
$$(\mathbf{a}, \mathbf{b}) \to \mathbf{a} \otimes \mathbf{b}$$

makes $T\mathbf{V}$ into an algebra called the *tensor algebra* of $\mathbf{V}$. (Theorem C.6.2(2) shows that we may assume that the product $\otimes$ is defined also when either r or s are 0.)

The tensor algebra is an example of what is called a *graded algebra* or *graded ring*, that is, an algebra or ring A that is a direct sum of additive subgroups $A_i$ with the property that the product of an element in $A_i$ and an element in $A_j$ lies in $A_{i+j}$.

Although we shall only be interested in tensor algebras, one usually generalizes the notation $T^k\mathbf{V}$ to allow for "mixed" tensors.

**Definition.** Let $\mathbf{V}$ be a vector space and $\mathbf{V}^*$ its dual. Define vector spaces $\mathbf{V}^r_s$ by

$$\mathbf{V}^0_0 = \mathbf{R}$$
$$\mathbf{V}^r_s = \underbrace{\mathbf{V} \otimes \mathbf{V} \otimes \cdots \otimes \mathbf{V}}_{r} \otimes \underbrace{\mathbf{V}^* \otimes \mathbf{V}^* \otimes \cdots \otimes \mathbf{V}^*}_{s}, \quad r + s > 0.$$

An element of $\mathbf{V}^r_s$, $r + s > 0$, is called a *tensor of type* $(r,s)$ or simply a *tensor* and is said to have *contravariant order r* and *covariant order s*. Elements of $\mathbf{V}^1_0$ are called *contravariant vectors* and elements of $\mathbf{V}^0_1$ are called *covariant vectors*.

Clearly, $T^r\mathbf{V}$ is the same as $\mathbf{V}^r_0$ and $T^s(\mathbf{V}^*)$ is the same as $\mathbf{V}^0_s$.

Next, let $\mathbf{U}_i$ and $\mathbf{V}_i$ be vector spaces and let $T_i : \mathbf{U}_i \to \mathbf{V}_i$ be linear transformations. Since the map

$$T_1 \times T_2 : \; \mathbf{U}_1 \times \mathbf{U}_2 \to \mathbf{V}_1 \otimes \mathbf{V}_2$$

defined by

$$(T_1 \times T_2)(\mathbf{u}_1, \mathbf{u}_2) = T_1(\mathbf{u}_1) \otimes T_2(\mathbf{u}_2)$$

is bilinear, there is unique linear transformation

$$T_1 \otimes T_2 : \; \mathbf{U}_1 \otimes \mathbf{U}_2 \to \mathbf{V}_1 \otimes \mathbf{V}_2.$$

so that

$$(T_1 \times T_2)(\mathbf{u}_1, \mathbf{u}_2) = (T_1 \otimes T_2)(\mathbf{u}_1 \otimes \mathbf{u}_2).$$

Because we do not have to worry about parenthesizing tensor products by Theorem C.6.2(3), this construction generalizes to produce a unique linear transformation

$$T_1 \otimes T_2 \otimes \cdots \otimes T_k : \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \cdots \otimes \mathbf{U}_k \to \mathbf{V}_1 \otimes \mathbf{V}_2 \otimes \cdots \otimes \mathbf{V}_k.$$

**Definition.** The map $T_1 \otimes T_2 \otimes \cdots \otimes T_k$ is called the *tensor product* of the maps $T_i$.

Now, although it is clear from the definition that tensor products have to do with multilinear maps, we want to describe a much more fundamental relationship between the two. See Theorem C.6.6 below. First of all we shall define a parallel algebra structure on multilinear maps.

**Definition.** If $f \in L^r(\mathbf{V})$ and $g \in L^s(\mathbf{V})$, then define $f \otimes g \in L^{r+s}(\mathbf{V})$ by

$$(f \otimes g)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_{r+s}) = f(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)g(\mathbf{v}_{r+1}, \ldots, \mathbf{v}_{r+s}), \tag{C.7}$$

for $\mathbf{v}_i \in \mathbf{V}$. The map $f \otimes g$ is called the *tensor product* of f and g.

**Notation.** Let $L(\mathbf{V})$ denote the direct sum of the vector spaces $L^k(\mathbf{V})$, $k \geq 0$.

**C.6.3. Theorem.** The tensor product operation $\otimes$ defined by equation (C.7) turns $L(\mathbf{V})$ into an algebra called the *algebra of real-valued multilinear maps* on $\mathbf{V}^k$.

**Proof.** This theorem is the analog of Theorem C.6.2(1–3). One can easily prove that $\otimes$ is associative and that the distributive laws hold.

**C.6.4. Theorem.** There is a unique vector space isomorphism

$$\psi : L^k(\mathbf{V}) \to (T^k\mathbf{V})^*, \tag{C.8a}$$

with the property that

$$g(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \psi(g)(\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_k) \tag{C.8b}$$

for all $g \in L^k(\mathbf{V})$ and $\mathbf{v}_i \in \mathbf{V}$.

**Proof.** This is an easy consequence of the universal factorization property of the tensor product.

**C.6.5. Theorem.** Let $\mathbf{U}$ and $\mathbf{V}$ be vector spaces.

(1) There is a unique isomorphism

$$\varphi : \mathbf{U}^* \otimes \mathbf{V}^* \to (\mathbf{U} \otimes \mathbf{V})^*$$

defined by the condition that

$$\varphi(\alpha \otimes \beta)(\mathbf{u} \otimes \mathbf{v}) = \alpha(\mathbf{u})\beta(\mathbf{v})$$

for all $\alpha \in \mathbf{U}^*$, $\beta \in \mathbf{V}^*$, $\mathbf{u} \in \mathbf{U}$, and $\mathbf{v} \in \mathbf{V}$.

(2) There is a unique isomorphism

$$\varphi: \ T^k(\mathbf{V}^*) \to (T^k\mathbf{V})^* \tag{C.9a}$$

defined by

$$\varphi(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k)(\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_k) = \alpha_1(\mathbf{v}_1)\alpha_2(\mathbf{v}_2)\cdots\alpha_k(\mathbf{v}_k) \tag{C.9b}$$

for all $\alpha_i \in \mathbf{V}^*$ and $\mathbf{v}_i \in \mathbf{V}$.

**Proof.** To prove part (1), note that the universal factorization property of tensor products implies that $\varphi$ is induced by the bilinear map

$$f: \ \mathbf{U}^* \times \mathbf{V}^* \to (\mathbf{U} \otimes \mathbf{V})^* \tag{C.10a}$$

defined by

$$f(\alpha,\beta)(\mathbf{u} \otimes \mathbf{v}) = \alpha(\mathbf{u})\beta(\mathbf{v}). \tag{C.10b}$$

To show that $\varphi$ is an isomorphism, show that the vectors $\varphi(\mathbf{u}_i^* \otimes \mathbf{v}_j^*)$ are linearly independent for some dual basis $\mathbf{u}_i^*$ for $\mathbf{U}^*$ and $\mathbf{v}_j^*$ for $\mathbf{V}^*$. See [AusM63] or [KobN63]. Part (2) is an easy generalization of part (1).

Because of Theorem C.6.5 we shall not distinguish between tensor products like $\mathbf{U}^* \otimes \mathbf{V}^*$ and $(\mathbf{U} \otimes \mathbf{V})^*$ whenever it is convenient. Note that another way to prove the existence of an isomorphism between these spaces is to show that $((\mathbf{U} \otimes \mathbf{V})^*, f)$ is a tensor product for $\mathbf{U}$ and $\mathbf{V}$, where f is the map in (C.10b). The same comment applies with regard to the existence of an isomorphism between $T^k(\mathbf{V}^*)$ and $(T^k\mathbf{V})^*$. We did not do this because for us it is convenient to be explicit about the formula (C.9b) for the isomorphism $\varphi$.

Using the isomorphism $\psi$ and $\varphi$ in Theorems C.6.4 and C.6.5(2), respectively, gives us an isomorphism

$$\psi^{-1}\circ\varphi: \ T^k(\mathbf{V}^*) \to L^k(\mathbf{V}). \tag{C.11}$$

**C.6.6. Theorem.** The isomorphisms in (C.11) induce an isomorphism of algebras

$$\Phi: \ T(\mathbf{V}^*) \to L(\mathbf{V}), \tag{C.12a}$$

where

$$\Phi(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k) = \alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k \tag{C.12b}$$

for all $\alpha_i \in \mathbf{V}^*$.

**Proof.** This is an easy exercise working through the definitions. Note that the left side of equation (C.12b) has a tensor product of simply "elements" or symbols $\alpha_i$ that happen to belong to $\mathbf{V}^*$ whereas the right side is a tensor product of **maps** as defined by (C.7).

Because multilinear maps are more intuitive than abstract tensor products, Theorem C.6.6 justifies our always treating $T(\mathbf{V}^*)$ as if it were $L(\mathbf{V})$. Notice, however, that we had to use the dual space $\mathbf{V}^*$ to make this identification.

Now we move on to a definition of exterior algebras. Let $S_k$ be the group of permutations of $\{1,2, \ldots ,k\}$. Any element $\sigma$ of $S_k$ induces a unique isomorphism

$$\sigma : T^k\mathbf{V} \to T^k\mathbf{V} \tag{C.13a}$$

satisfying

$$\sigma(\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_k) = \mathbf{v}_{\sigma(1)} \otimes \mathbf{v}_{\sigma(2)} \otimes \cdots \otimes \mathbf{v}_{\sigma(k)}, \ \ \mathbf{v}_i \in \mathbf{V}. \tag{C.13b}$$

**Definition.** A tensor $\alpha$ in $T^k\mathbf{V}$ is said to be *alternating* if $\sigma(\alpha) = \text{sign}(\sigma)\alpha$, for all $\sigma \in S_k$. A linear transformation

$$T : T^k\mathbf{V} \to \mathbf{W}$$

is said to be *alternating* if $T \circ \sigma = \text{sign}(\sigma)T$, for all $\sigma \in S_k$.

Now, by Theorem C.6.6 we can identify $T^k(\mathbf{V}^*)$ with $(T^k\mathbf{V})^*$.

**C.6.7. Theorem.** A tensor in $T^k(\mathbf{V}^*)$ is alternating if and only if it corresponds to an alternating linear transformation in $(T^k\mathbf{V})^*$ under the natural isomorphism $\varphi$ defined by Theorem C.6.5(2).

**Proof.** See [AusM63].

**Definition.** The *alternation map* $\text{Alt}: T^k\mathbf{V} \to T^k\mathbf{V}$ is defined by

$$\begin{aligned} \text{Alt} &= \frac{1}{k!}\sum_{\sigma \in S_k} (\text{sign}\ \sigma)\sigma, \qquad \text{if } k \geq 1, \\ &= \text{the identity map}, \quad \text{if } k = 0. \end{aligned}$$

**C.6.8. Theorem.**

(1) The alternation map $\text{Alt}: T^k\mathbf{V} \to T^k\mathbf{V}$ is a linear transformation.
(2) If $\alpha \in T^k\mathbf{V}$, then $\text{Alt}(\alpha)$ is an alternating tensor.
(3) If $\alpha \in T^k\mathbf{V}$ is an alternating tensor, then $\text{Alt}(\alpha) = \alpha$.

**Proof.** This is an easy exercise.

Theorem C.6.8 shows that $\text{Alt}^2 = \text{Alt}$, so that Alt is a projection of $T^k\mathbf{V}$ onto the subspace of alternating tensors.

**C.6.9. Theorem.** Let $T: T^k\mathbf{V} \to \mathbf{W}$ be a linear transformation. The map T is alternating if and only if T(ker Alt) = **0**.

**Proof.** See [AusM63].

**Definition.** Let **V** be a vector space. Define the *k-fold exterior product* of **V**, denoted by $E^k\mathbf{V}$, to be the subspace of alternating tensors in $T^k\mathbf{V}$, that is, using Theorem C.6.8(2),

$$E^0\mathbf{V} = T^0\mathbf{V} = \mathbf{R},$$
$$E^k\mathbf{V} = \mathrm{Alt}(T^k\mathbf{V}) \subseteq T^k\mathbf{V},\ k \geq 1,$$

Clearly, $E^1\mathbf{V} = \mathbf{V}$.

Next, we would like to define a product for alternating tensors that maps alternating tensors to alternating tensors. The tensor product of two alternating tensors is unfortunately not always an alternating tensor, but all we have to do is project back into the set of alternating tensors using the Alt map.

**Definition.** Define a map

$$\wedge:\ E^r(\mathbf{V}) \times E^s(\mathbf{V}) \to E^{r+s}(\mathbf{V})$$

by

$$\omega \wedge \eta = \frac{(r+s)!}{r!s!}\,\mathrm{Alt}(\omega \otimes \eta). \tag{C.14}$$

The map $\wedge$ is called the *exterior* or *wedge product*.

The ugly factorials are added here in order to avoid them in other places, such as in the definition of the volume element for differential forms.

**C.6.10. Theorem.** Let **V** be an n-dimensional vector space.

(1) The exterior product for **V** is bilinear, that is,

$$(\omega_1 + \omega_2) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$$
$$\omega \wedge (\eta_1 \wedge \eta_2) = \omega \wedge \eta_1 + \omega \wedge \eta_2$$
$$a\omega \wedge \eta = \omega \wedge a\eta = a(\omega \wedge \eta)$$

for $\omega_i, \omega \in E^r(\mathbf{V})$, $\eta_i, \eta \in E^s(\mathbf{V})$, and $a \in \mathbf{R}$.

(2) If $\omega \in E^r(\mathbf{V})$, $\eta \in E^s(\mathbf{V})$, and $\theta \in E^t(\mathbf{V})$, then

$$(\omega \wedge \eta) \wedge \theta = \omega \wedge (\eta \wedge \theta) = \frac{(r+s+t)!}{r!s!t!}\,\mathrm{Alt}(\omega \otimes \eta \otimes \theta).$$

In particular, the exterior product is associative and

$$\alpha_1 \wedge \alpha_2 \wedge \cdots \alpha_k = k!\,\mathrm{Alt}(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_k)$$

for $\alpha_i \in E^1(\mathbf{V})$.

(3) If $\omega \in E^r(\mathbf{V})$ and $\eta \in E^s(\mathbf{V})$, then

$$\omega \wedge \eta = (-1)^{rs} \eta \wedge \omega.$$

In particular, if $\alpha, \beta \in E^1(\mathbf{V})$, then $\alpha \wedge \beta = -\beta \wedge \alpha$ and $\alpha \wedge \alpha = 0$.

(4) If $\sigma \in S_k$ and $\alpha_i \in E^1(\mathbf{V})$, then

$$\alpha_{\sigma(1)} \wedge \alpha_{\sigma(2)} \wedge \cdots \wedge \alpha_{\sigma(k)} = (\mathrm{sign}\,\sigma)\,\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k.$$

(5) Let $\alpha_1, \alpha_2, \ldots,$ and $\alpha_n$ form a basis for $E^1(\mathbf{V})$. If $1 \le k \le n$, then the set of all

$$\alpha_{i_1} \wedge \alpha_{i_2} \wedge \cdots \wedge \alpha_{i_k},\ 1 \le i_1 < i_2 < \cdots < i_k \le n,$$

is a basis for $E^k(\mathbf{V})$. It follows that $E^k(\mathbf{V})$ has dimension $\binom{n}{k}$ for $0 \le k \le n$.

(6) If $k > n$, then $E^k(\mathbf{V}) = 0$.

**Proof.** The proofs in [Spiv65] for alternating multilinear maps readily translate to our situation here.

**Definition.** Let $E\mathbf{V}$ denote the direct sum of the vector spaces $E^k(\mathbf{V})$, $k \ge 0$. The exterior product $\wedge$ makes $E\mathbf{V}$ into an algebra called the *exterior algebra* or *Grassmann algebra* of $\mathbf{V}$.

**C.6.11. Theorem.** Let $\mathbf{V}$ be a vector space and $\mathbf{V}^*$ its dual space. For $k \ge 1$, there is a unique isomorphism

$$\Phi:\ E^k(\mathbf{V}^*) \to (E^k\mathbf{V})^*$$

such that

$$\Phi(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_k) = \det(\alpha_i(\mathbf{v}_j)),$$

for $\alpha_i \in \mathbf{V}^*$ and $\mathbf{v}_j \in \mathbf{V} = E^1(\mathbf{V})$.

**Proof.** See [AusM63].

**C.6.12. Theorem.** Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces. A linear transformation $T: \mathbf{V} \to \mathbf{W}$ induces a unique linear transformation

$$E^kT:\ E^k\mathbf{V} \to E^k\mathbf{W}$$

such that

$$E^kT(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_k) = T(\mathbf{v}_1) \wedge T(\mathbf{v}_2) \wedge \cdots \wedge T(\mathbf{v}_k)$$

for $\mathbf{v}_i \in \mathbf{V}$.

**Proof.** See [AusM63].

**Definition.** The map $E^kT$ in Theorem C.6.12 is called the *k-fold exterior product* of T.

Theorem C.6.6 showed the relationship between tensor algebras and multilinear map algebras. Now we want to show how alternating tensors are related to special types of multilinear maps in a similar way.

Given a permutation $\sigma \in S_k$, define

$$\sigma : \mathbf{V}^k \to \mathbf{V}^k$$

by

$$\sigma(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = (\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \ldots, \mathbf{v}_{\sigma(k)}).$$

**Definition.** Let **V** and **W** be vector spaces. A multilinear map

$$T : \mathbf{V}^k \to \mathbf{W}$$

is said to be *alternating* if $T \circ \sigma = (\text{sign } \sigma)T$ for all $\sigma \in S_k$. The set of such alternating maps will be denoted by $\Lambda^k(\mathbf{V};\mathbf{W})$. If $\mathbf{W} = \mathbf{R}$, then $\Lambda^k(\mathbf{V};\mathbf{W})$ will be abbreviated to $\Lambda^k(\mathbf{V})$.

It is easy to show that $\Lambda^k(\mathbf{V};\mathbf{W})$ is a vector space. By definition,

$$\Lambda^k(\mathbf{V};\mathbf{W}) \subseteq L^k(\mathbf{V};\mathbf{W}) \quad \text{and} \quad \Lambda^k(\mathbf{V}) \subseteq L^k(\mathbf{V}).$$

The well-known properties of the determinant function lead to the standard example of an alternating mulilinear map.

**Definition.** The map

$$\det : \mathbf{R}^n \to \mathbf{R}$$

defined by

$$\det((a_{11}, a_{12}, \ldots, a_{1n}), (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, (a_{nl}, a_{n2}, \ldots, a_{nn})) = \det(a_{ij})$$

is called the *determinant map* of $\mathbf{R}^n$.

Clearly, $\det \in \Lambda^n(\mathbf{R}^n)$.

The next theorem shows that, like the tensor product, the k-fold exterior product could have been defined in terms of a *universal factorization property* with respect to alternating multilinear maps. Let

$$\lambda : \mathbf{V}^k \to E^k\mathbf{V} \tag{C.15}$$

be the composite of the maps

$$\mathbf{V}^k \xrightarrow{\otimes} T^k\mathbf{V} \xrightarrow{\text{Alt}} E^k\mathbf{V}.$$

**C.6.13. Theorem.** Let $g : \mathbf{V}^k \to \mathbf{W}$ be an alternating multilinear map. Then there is a unique linear transformation $h : E^k\mathbf{V} \to \mathbf{W}$ so that $g = h \circ \lambda$. In fact, the map

$$g \to h$$

defines an isomorphism between alternating multilinear maps on $\mathbf{V}^k$ and linear maps on $E^k\mathbf{V}$. As a special case we get an isomorphism

$$\psi : \Lambda^k(\mathbf{V}) \to (E^k\mathbf{V})^*,$$

where

$$g(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \psi(g)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_k) \tag{C.16}$$

for all $g \in L^k(\mathbf{V})$ and $\mathbf{v}_i \in \mathbf{V}$.

**Proof.** See Figure C.5. The theorem is an easy consequence of Theorem C.6.9. For a proof see [AusM63].

Theorem C.6.1 and C.6.2 remain true if we replace "multilinear map" with "alternating multilinear map".

Alternating multilinear maps admit a product very much like the alternating tensor product.

**Definition.** The *alternation map*

$$\text{Alt} : L^k(\mathbf{V}) \to L^k(\mathbf{V})$$

is defined by

$$\begin{aligned} \text{Alt}(T) &= \frac{1}{k!} \sum_{\sigma \in S_k} (\text{sign}\, \sigma)\, T \circ \sigma, \text{ if } k \geq 1, \\ &= T, \text{ if } k = 0. \end{aligned}$$

$$\begin{tikzcd} V^k \arrow[rr, "\lambda"] \arrow[dr, "g"'] & & E^kV \arrow[dl, dashed, "h"] \\ & W & \end{tikzcd}$$

**Figure C.5.** The universal property of the space $E^k\mathbf{V}$.

**C.6.14. Theorem.**

(1) The alternation map $\mathrm{Alt}: L^k(\mathbf{V}) \to L^k(\mathbf{V})$ is a linear transformation.
(2) If $\alpha \in L^k(\mathbf{V})$, then $\mathrm{Alt}(\alpha)$ is an alternating multilinear map.
(3) If $\alpha \in L^k(\mathbf{V})$ is an alternating multilinear, then $\mathrm{Alt}(\alpha) = \alpha$.

**Proof.** This is easy and proved like Theorem C.6.8.

Theorem C.6.14 implies that

$$\Lambda^k(\mathbf{V}) = \mathrm{Alt}\left(L^k(\mathbf{V})\right)$$

and is a direct summand of $L^k(\mathbf{V})$. We can define a product, called the *exterior* or *wedge product*,

$$\wedge: \ \Lambda^r(\mathbf{V}) \times \Lambda^s(\mathbf{V}) \to \Lambda^{r+s}(\mathbf{V}) \tag{C.17}$$

by the same formula (C.15) that we used for the exterior algebra. Theorem C.6.10 holds verbatim for alternating multilinear maps. This is exactly how Spivak ([Spiv65]) develops the exterior algebra.

**Definition.** An element of $\Lambda^k(\mathbf{V})$ is called an *exterior k-form* on $\mathbf{V}$. Let $\Lambda\mathbf{V}$ denote the direct sum of the vector spaces $\Lambda^k(\mathbf{V})$, $k \geq 0$. The exterior product $\wedge$ makes $\Lambda\mathbf{V}$ into an algebra called the *algebra of exterior forms* on $\mathbf{V}$.

Finally, we already know from Theorem C.6.11 and Theorem C.6.13 that there are natural isomorphisms

$$E^k(\mathbf{V}^*) \to \left(E^k\mathbf{V}\right)^* \leftarrow \Lambda^k(\mathbf{V}), \tag{C.18}$$

so that $E^k(\mathbf{V}^*)$ and $\Lambda^k(\mathbf{V})$ are isomorphic.

**C.6.15. Theorem.** The isomorphisms in (C.18) induce an isomorphism of algebras

$$\Phi: \ E(\mathbf{V}^*) \to \Lambda(\mathbf{V}), \tag{C.19a}$$

where

$$\Phi(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k) = \alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k \tag{C.19b}$$

for all $\alpha_i \in \mathbf{V}^*$.

**Proof.** This is another easy exercise working through the definitions. Note that the left side of Equation (C.19b) has an exterior product of simply "elements" or symbols $\alpha_i$ that happen to belong to $\mathbf{V}^*$ whereas the right side is an exterior product of **maps** as defined by (C.17).

Theorems C.6.6 and C.6.15 can be summarized compactly by saying that we have a commutative diagram

$$\begin{array}{ccc} T(\mathbf{V}^*) & \xrightarrow{\Phi} & L(\mathbf{V}) \\ \iota \ \cup & & \cup \ \iota \\ E(\mathbf{V}^*) & \xrightarrow{\Phi} & \Lambda(\mathbf{V}) \end{array}$$

of inclusion maps $\iota$ and algebra isomorphisms $\Phi$.

We finish this section with a few more facts about $\Lambda(\mathbf{V})$.

**C.6.16. Theorem.** Let $\mathbf{V}$ be an n-dimensional vector space and $\mathbf{V}^*$ its dual. If $\alpha_i \in \mathbf{V}^*$, then the element $\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_k \in \Lambda^k(\mathbf{V})$ satisfies

$$(\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_k)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \det(\alpha_i(\mathbf{v}_j))$$

for all $\mathbf{v}_j \in \mathbf{V}$.

**Proof.** This is a corollary of Theorem C.6.11.

Finally, if $T: \mathbf{V} \to \mathbf{W}$ is a linear transformation, then the induced map

$$T^*: \ \mathbf{W}^* \to \mathbf{V}^*$$

on dual spaces defines

$$E^k(T^*): \ E^k(\mathbf{W}^*) \to E^k(\mathbf{V}^*).$$

**C.6.17. Theorem.** Under our identifications, the map $E^k(T^*)$ induces the map

$$T^*: \ \Lambda^k(\mathbf{W}) \to \Lambda^k(\mathbf{V})$$

defined by

$$T^*(\alpha)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \alpha(T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_k)), \ \alpha \in \Lambda^k(\mathbf{W}), \ \mathbf{v}_i \in \mathbf{V}.$$

**Proof.** One simply has to carefully work through all the appropriate identifications.

It is the tensor algebra $T\mathbf{V}^*$ and the exterior algebra $E\mathbf{V}^*$ that are the most interesting because they formalize the algebra of multilinear maps and the algebra of alternating multilinear maps, respectively, which are needed for defining differential forms.

We finish with one final observation. One can use the exterior algebra to define determinants. For example, let $\mathbf{V}$ be an n-dimensional vector space and

$$T: \ \mathbf{V} \to \mathbf{V}$$

a linear transformation. We know that the vector space $E^n\mathbf{V}$ has dimension 1. Therefore, the linear transformation

$$E^nT: \ E^n\mathbf{V} \to E^n\mathbf{V}$$

has the form

$$E^n T(\alpha) = d\alpha, \tag{C.20}$$

for some unique constant d. This constant is of course the determinant of T and one could make equation (C.20) the basis of a definition of the determinant. This can also be used to define the determinant of a matrix. Advanced books on algebra that discuss multilinear maps often use this approach to determinants because, having developed the machinery of tensors one then gets many of the properties of determinants for almost "free."

APPENDIX D

# Basic Calculus and Analysis

We assume that the reader is familiar with limits, continuity, the derivative, and the Riemann integral of real-valued functions of a real variable.

## D.1 Miscellaneous Facts

We start by recalling some standard notation.

**Definition.** Let a, b $\in$ **R**. Define the *closed interval* [a,b], the *open interval* (a,b), and the *half-open intervals* [a,b) and (a,b] to be the sets

$$[a,b] = \{t \mid a \le t \le b\},\ (a,b) = \{t \mid a < t < b\},$$

$$[a,b) = \{t \mid a \le t < b\}, \quad \text{and} \quad (a,b] = \{t \mid a < t \le b\}.$$

Note that if b < a, then [a,b] = (a,b) = $\phi$.

**Definition.** Let **X** be a nonempty set of real numbers. A *lower bound* for **X** is a number c so that c $\le$ x for all x $\in$ **X**. If **X** has a lower bound, then the *infimum* or *greatest lower bound* for **X**, denoted by inf **X**, is the largest element in the set of lower bounds. An *upper bound* for **X** is a number c so that c $\ge$ x for all x $\in$ **X**. If **X** has an upper bound, then the *supremum* or *least upper bound* for **X**, denoted by sub **X**, is the smallest element in the set of upper bounds. (The completeness of the real numbers guarantees that the largest and smallest elements exist in our cases.)

**Definition.** A function defined on an interval is said to be *monotonic* if it is either increasing or decreasing on that interval.

Next, we recall some limit notation. The reader is assumed to know about (two-sided) limits, but sometimes there is a need to talk about one-sided limits, such as at end of intervals. The notation used in that case will be the following: For a function f, the *right-handed* and *left-handed* limit at a point a denoted by $f(a^+)$ and $f(a^-)$, respectively, are defined by

$$f(a^+) = \lim_{x \to a, x \geq a} f(x) \quad \text{and} \quad f(a^-) = \lim_{x \to a, x \leq a} f(x).$$

**D.1.1. Theorem.** (The Intermediate Value Theorem) Let $f:[a,b] \to \mathbf{R}$ be a continuous function. Assume that $c \in [f(a),f(b)]$. Then there exists an $\alpha \in [a,b]$ so that $f(\alpha) = c$.

**Proof.** See [Buck78].

**D.1.2. Theorem.** (The Mean Value Theorem) Let $f:[a,b] \to \mathbf{R}$ be continuous and assume that f is differentiable on (a,b). Then there exists an $\alpha \in (a,b)$ so that

$$f(b) - f(a) = (b - a)f'(\alpha).$$

**Proof.** See [Buck78].

**D.1.3. Theorem.** (The Leibnitz Formula) Suppose that $h(x) = f(x)g(x)$, where f(x) and g(x) are n-times differentiable function. Then the product rule for the derivative generalizes to

$$h^{(n)}(x) = \sum_{i=0}^{n} \binom{n}{i} f^{(i)}(x) g^{(n-i)}(x).$$

**Proof.** Use induction.

**Definition.** Let f be a real-valued function defined on all of $\mathbf{R}$ or an open interval (a,b). The function f is said to be of *class* $C^k$ if all the derivatives of f exist **and** are continuous up to and including order k. If f is of class $C^k$ for all k, then we say that f is of *class* $C^\infty$.

**Definition.** A *partition* of an interval [a,b] is a sequence

$$P = (t_0, t_1, \ldots, t_k), \quad \text{where } a = t_0 \leq t_1 \leq \cdots \leq t_k = b.$$

Each interval $[t_i,t_{i+1}]$ is called a *subinterval* of P. The *norm* of the partition P, denoted by $|P|$, is defined by

$$|P| = \max \{|t_i - t_{i-1}| \mid i = 1, 2, \ldots, k\}.$$

A *refinement* of the partition P is a partition $P' = (s_0,s_1, \ldots ,s_m)$ of [a,b] with the property that $\{t_0,t_1, \ldots ,t_k\} \subseteq \{s_0,s_1, \ldots ,s_m\}$.

When it comes to integration, we assume that the reader is familiar with the Riemann integral, but we recall a few basic definitions and facts. Given a bounded function $f:[a,b] \to \mathbf{R}$ the standard definition of the Riemann integral is in terms of a limit of sums. More precisely, for each partition $P = (x_0,x_1, \ldots ,x_n)$ of [a,b] we look at sums of the form

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}),$$

where $\xi_i \in [x_{i-1}, x_i]$. If these sums converge as the norm of the partition goes to zero, then their limit is called the *Riemann integral* of f over [a,b] and denoted by

$$\int_a^b f.$$

**Definition.** A function F is an *antiderivative* of a function f if $F'(x) = f(x)$ for all x in their common domain.

**D.1.4. Theorem.** (The Fundamental Theorem of Calculus) Let $f:[a,b] \to \mathbf{R}$ be a continuous function.

(1) The function f has an antiderivative.
(2) If F is any antiderivative of f, then

$$\int_a^b f(x)dx = F(b) - F(a).$$

**Proof.** See [Buck78].

**D.1.5. Theorem.** (The Change of Variable Theorem) Let $\phi:[\alpha,\beta] \to \mathbf{R}$ be a continuously differentiable function on $[\alpha,\beta]$ and let $\phi(\alpha) = a$ and $\phi(\beta) = b$. If f is a continuous function on $\phi([\alpha,\beta])$, then

$$\int_a^b f(x)dx = \int_\alpha^\beta f(\phi(u))\phi'(u)du.$$

**Proof.** See [Buck78].

**Definition.** A function f is said to be *absolutely integrable* if $|f|$ is integrable.

Finally, there are times when one needs to consider integrals over unbounded regions. The definitions for such integrals, also called *improper integrals*, are fairly straightforward. They are defined as limits of integrals over finite domains, assuming that the limits exist. More precisely, in the one variable case one defines

$$\int_a^\infty f = \lim_{b\to\infty}\int_a^b f \quad \text{and} \quad \int_{-\infty}^b f = \lim_{a\to-\infty}\int_a^b f$$

with

$$\int_{-\infty}^\infty f = \int_0^\infty f + \int_{-\infty}^0 f.$$

We finish this section with two improper integrals whose values are worth knowing on occasion.

**D.1.6. Theorem.**

$$\int_0^\infty \sin(x^2)dx = \int_0^\infty \cos(x^2)dx = \sqrt{\frac{\pi}{8}}. \tag{D.1}$$

**Proof.** See [Buck78].

**Definition.** The integrals in (D.1) are called *Fresnel integrals*.

## D.2 Series

This section reviews some basic facts about series, in particular Taylor polynomials and series.

**Definition.** A series

$$\sum_{n=0}^{\infty} a_n \tag{D.2}$$

is said to *converge* to the sum A if the sequence of partial sums

$$\sum_{n=0}^{k} a_n$$

converges to A as k goes to infinity, otherwise, it is said to *diverge*. If the series (D.2) converges, but the series

$$\sum_{n=0}^{\infty} |a_n| \tag{D.3}$$

diverges, (D.2) is called a *conditionally convergent* series. If (D.3) converges, then (D.2) is called an *absolutely convergent* series.

**D.2.1. Theorem.** Every absolutely convergent series converges.

**Proof.** See [Buck78].

**Definition.** Series of the form

$$\sum_{n=0}^{\infty} a_n x^n \tag{D.4}$$

or

$$\sum_{n=0}^{\infty} a_n (x-c)^n \tag{D.5}$$

are called *power series* in x or x – c, respectively.

**D.2.2. Theorem.** For every power series of the form (D.4) there is an R, $0 \le R \le \infty$, so that the series converges absolutely for all x, $|x| < R$, and diverges for all x, $R < |x|$. One can compute R with the formulas

$$\frac{1}{R} = \limsup_{n\to\infty} |a_n|^{1/n} = \sup\left\{|a_n|^{1/n}\right\} \tag{D.6a}$$

or

$$\frac{1}{R} = \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} \tag{D.6b}$$

provided that the limits exist.

**Proof.** See [Buck78].

**Definition.** The number R in formulas (D.6) is called the *radius of convergence* of the series (D.4).

Clearly, if R is the radius of convergence of the power series (D.4), then (D.5) will converge for all x satisfying $|x - c| < R$, and diverge for all x satisfying $|x - R| > R$. Therefore, the points at which a power series converges is an open interval together with possibly its endpoints. The endpoints of the interval typically have to be checked separately for convergence or divergence.

**Definition.** The interval of numbers at which the power series (D.4) or (D.5) converges is called its *interval of convergence*.

**D.2.3. Theorem.** Let R be the radius of convergence of the series defined by (D.4). The function it defines is differentiable for all x, $|x| < R$. Its derivative can be obtained by termwise differentiation and its radius of convergence is again R.

**Proof.** See [Buck78].

**Definition.** Let $f:(a,b) \to \mathbf{R}$ be of class $C^k$. Let $x_0 \in (a,b)$. The polynomial

$$f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k$$

is called the *Taylor polynomial of f of degree k at $x_0$*.

**D.2.4. Theorem.** (The Taylor Polynomial Theorem) Let $f:(a,b) \to \mathbf{R}$ be of class $C^{k+1}$ and let $c \in (a,b)$. Then for any $x \in (a,b)$ there is an $\alpha \in [c,x]$ such that

$$f(x) = f(c) + f'(c)(x - c) + \cdots + \frac{1}{k!} f^{(k)}(c)(x - c)^k + \frac{1}{(k+1)!} f^{(k+1)}(\alpha)(x - c)^{k+1}.$$

**Proof.** See [Buck78].

**Definition.** Let $f:(a,b) \to \mathbf{R}$ be of class $C^\infty$. Let $P_c(x)$ be the Taylor polynomial of f of degree k at $c \in (a,b)$. Let $R_k(x) = f(x) - P_c(x)$. The function f is said to be *analytic* at c if there is an open interval **I** in (a,b) containing c such that

$$\lim_{k\to\infty} R_k(x) = 0$$

for all $x \in \mathbf{I}$.

Analytic functions are properly discussed in complex analysis. See Appendix E.

**Definition.** Let $f:(a,b) \to \mathbf{R}$ be of class $C^\infty$ and let $c \in (a,b)$. The power series

$$\sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(c)(x-c)^n$$

is called the *Taylor series for f about c*.

## D.3 Differential Equations

For the sake of completeness we shall state the two theorems about solutions to differential equations that are needed in this book. This section uses a few concepts from Sections 4.2 and 4.3. Note that the *order* of an (ordinary) differential equation is the order of the highest derivative appearing in the equation.

**Problem 1:** Let $\mathbf{D}$ be a connected open subset of $\mathbf{R}^{n+1}$ and assume that $f_1, f_2, \ldots, f_n : \mathbf{D} \to \mathbf{R}$ are continuous functions. We want to find an open interval (a,b) and functions $\varphi_i:(a,b) \to \mathbf{R}$, so that

(1) $(t,\varphi_1(t),\varphi_2(t), \ldots ,\varphi_n(t)) \in \mathbf{D}$, and
(2) $\varphi_i'(t) = f_i(t,\varphi_1(t),\varphi_2(t), \ldots ,\varphi_n(t))$.

for $t \in (a,b)$.

The equations in Problem 1 are called a *system of n ordinary differential equations of the first order* and the differentiable functions $\varphi_i(t)$, if they exist, are called *solutions* to the system.

**D.3.1. Theorem.** Let $(t_0,x_1,x_2, \ldots ,x_n) \in \mathbf{D}$. Then there exists an $\varepsilon > 0$ and unique continuously differentiable functions $\varphi_i:(t_0 - \varepsilon,t_0 + \varepsilon) \to \mathbf{R}$ that are solutions to Problem 1 and satisfy $\varphi_i(t_0) = x_i$.

**Proof.** See [CodL55].

**Problem 2:** Let $\mathbf{D}$ be a connected open subset of $\mathbf{R}^{n+1}$ and assume that $f:\mathbf{D} \to \mathbf{R}$ is a continuous function. We want to find an open interval (a,b) and a function $\varphi:(a,b) \to \mathbf{R}$, so that

(1) $(t,\varphi(t),\varphi'(t), \ldots ,\varphi^{(n-1)}(t)) \in \mathbf{D}$, and
(2) $\varphi^{(n)}(t) = f(t,\varphi(t),\varphi'(t), \ldots ,\varphi^{(n-1)}(t))$.

for $t \in (a,b)$.

The equation in Problem 2 is called the *nth-order differential equation* associated to the function f and the n times differentiable function $\varphi(t)$, if it exists, is called a *solution* to the equation.

**D.3.2. Theorem.** Let $(t_0, x_1, x_2, \ldots, x_n) \in \mathbf{D}$. Then there exists an $\varepsilon > 0$ and unique n times continuously differentiable function $\varphi: (t_0 - \varepsilon, t_0 + \varepsilon) \to \mathbf{R}$ that is the solution to Problem 2 and satisfies $\varphi^{(i)}(t_0) = x_{i+1}$, $0 \le i < n$.

**Proof.** To prove the theorem one reduces Problem 2 to Problem 1 by introducing new function $\varphi_i$ and solving

$$\begin{aligned}
\varphi_1'(t) &= \varphi_2(t)\\
\varphi_2'(t) &= \varphi_3(t)\\
\vdots \qquad & \quad \vdots\\
\varphi_{n-1}'(t) &= \varphi_n(t)\\
\varphi_n'(t) &= f(t, \varphi_1(t), \varphi_2(t), \ldots, \varphi_n(t)).
\end{aligned}$$

See [CodL55].

The values $t_0$, $x_1$, $x_2$, . . . , and $x_n$ in Theorem D.3.1 and D.3.2 are called *initial conditions*. The theorems can be rephrased as saying that initial conditions specify a unique **local** solution. An interesting question is whether these local solutions extend to global solutions. In two important special cases this is indeed the case.

Consider the *linear system of differential equations of the first order*

$$\begin{aligned}
y_1'(x) &= a_{11}(x)y_1 + a_{12}(x)y_2 + \cdots + a_{1n}(x)y_n,\\
y_2'(x) &= a_{21}(x)y_1 + a_{22}(x)y_2 + \cdots + a_{2n}(x)y_n,\\
\vdots \qquad & \quad \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots\\
y_n'(x) &= a_{n1}(x)y_1 + a_{n2}(x)y_2 + \cdots + a_{nn}(x)y_n.
\end{aligned} \tag{D.7}$$

Assume that the functions $a_{ij}(x)$ are continuous over some interval $\mathbf{X}$ which could be open, closed, or all of $\mathbf{R}$.

**D.3.3. Theorem.** Let $x_0 \in \mathbf{X}$ and let $c_0$, $c_1$, . . . , $c_{n-1}$ be arbitrary real numbers. There exist unique functions $y_i(x)$ defined on $\mathbf{X}$ with continuous derivatives satisfying Equations (D.7) and the conditions

$$y_1(x_0) = c_1, \quad y_2(x_0) = c_2, \ \ldots, \quad y_n(x_0) = c_n.$$

**Proof.** See [CodL55].

**Definition.** A *linear differential equation* is any equation of the form

$$a_0(x)y^{(n)}(x) + a_1(x)y^{(n-1)}(x) + \cdots + a_{n-1}(x)y'(x) + a_n(x)y(x) = f(x), \tag{D.8}$$

where $a_i(x)$, $y(x)$, $f(x)$ are functions and $y^{(i)}(x)$ denotes the ith derivative of $y(x)$.

We shall assume that the functions $a_i(x)$ and $f(x)$ in Equation (D.8) are continuous over some interval $\mathbf{X}$ which could be open, closed, or all of $\mathbf{R}$ and that $a_0(x) \neq 0$ for $x \in \mathbf{X}$.

**D.3.4. Theorem.** Let $x_0 \in \mathbf{X}$ and let $c_0, c_1, \ldots, c_{n-1}$ be arbitrary real numbers. There is a unique function y(x) defined on **X** with continuous derivatives up to order n that satisfies Equation (D.8) and the conditions

$$y(x_0) = c_0,\ y'(x_0) = c_1, \ldots, y^{(n-1)}(x_0) = c_{n-1}.$$

**Proof.** See [CodL55].

## D.4 The Lebesgue Integral

The Riemann integral is quite adequate for most tasks. Certainly, for all the functions dealt with explicitly in this book and the functions the reader might think about, the reader will not be wrong if he/she treat their integrals as Riemann integrals. This integral does has some drawbacks however that another integral called the Lebesgue integral does not have. The Lebesgue integral is more general and is better for some mathematical topics for technical reasons because one gets nicer and more complete results. Fortunately, the two integrals agree on most of the functions that are of practical interest. Nevertheless, there are some differences, in particular when one is integrating over unbounded intervals.

The specific reason for bringing up the Lebesgue integral in this book is its connection with the mathematics behind understanding the aliasing problem in computer graphics. See [AgoM05]. This problem is one of the first problems that one encounters in computer graphics and is caused by the fact that one is trying to display continuous objects in a discrete way. Alleviating this problem involves understanding some fairly fancy mathematics such as Fourier series and Fourier transforms. See Chapter 21 of [AgoM05]. It turns out that the Lebesgue integral would make it easier and clearer to state some of the definitions and theorems in that chapter carefully and correctly. It is a natural integral to use in the area of signal processing and the reader may see references to it in the context of digital image processing.

We cannot give the definition of the Lebesgue integral here because that would entail the discussion of yet another topic called measure theory. Instead we refer the interested reader to textbooks which cover this subject such as [Berb66], [Nata61], and [Spie69]. This is the one case in this book where a term is used without giving its definition. Our only goal is state a few facts that will at least show the reader the close relationship between Riemann and Lebesgue integration for functions of one variable. Similar results hold for integrals of functions of more variables.

**D.4.1. Theorem.** Let $f:[a,b] \to \mathbf{R}$.

(1) If f is Riemann integrable, then it is Lebesgue integrable.
(2) If f is Lebesgue integrable, then $|f|$ is Lebesgue integrable. (This is not true for the Riemann integral because there are functions f with the property that f is Riemann integrable but $|f|$ is not.)
(3) If $|f|$ if Riemann integrable, then f is both Riemann and Lebesgue integrable and both integrals are equal.

**Proof.** See the references mentioned above.

Theorem D.4.1(3) also holds in the unbounded case. The next example shows that, unlike the Lebesgue integral, there are functions f with the property that f is Riemann integrable but |f| is not.

**D.4.2. Example.** The Riemann integral

$$\int_0^\infty \frac{\sin x}{x} dx$$

exists, but the Riemann integral

$$\int_0^\infty \frac{|\sin x|}{|x|} dx$$

diverges. See [Spie69].

Finally, the Riemann integral is only defined if the function is bounded. This is not a requirement for the Lebesgue integral.

APPENDIX E

# Basic Complex Analysis

## E.1 Basic Facts

This appendix summarizes those facts about the complex numbers $\mathbf{C}$ and complex analysis that are needed in other parts of the book. A standard general reference on complex analysis is [Ahlf66], but [Need98] is also a book that the author would recommend.

We can write a complex number $\mathbf{z} \in \mathbf{C}$ in the form

$$\mathbf{z} = r(\cos\theta + \mathbf{i}\sin\theta).$$

**Definition.** This representation is called the *polar form* representation of the complex number $\mathbf{z}$ and the function $\arg(\mathbf{z})$ defined by $\arg(\mathbf{z}) = \theta$ is called the *argument* function.

Note that $\arg(\mathbf{z})$ is a multiple-valued function. Also,

$$\arg(\mathbf{z}_1\mathbf{z}_2) = \arg(\mathbf{z}_1) + \arg(\mathbf{z}_2).$$

**Definition.** The *extended complex plane* is obtained by adding one extra point to $\mathbf{C}$. This point is denoted by $\infty$ and called the *point at infinity*. (Topologically, the extended complex plane is the one-point compactification of $\mathbf{C}$. See Section 5.5.)

We identify the extended complex plane with the unit sphere $\mathbf{S}^2$ in $\mathbf{R}^3$ using the stereographic projection.

**Definition.** The sphere $\mathbf{S}^2$ along with the complex structure induced by the stereographic projection is called the *Riemann sphere*.

**Definition.** A point $(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n) \in \mathbf{C}^n$ is called a *real point* if all of its coordinates $\mathbf{z}_i$ are real.

Finally, since $\mathbf{C}$ and $\mathbf{R}^2$ are the same sets and we have a notion of limit and continuity for maps defined on $\mathbf{R}^2$, these notions carry over to maps $f: \mathbf{C} \to \mathbf{C}$.

## E.2 Analytic Functions

**Definition.** Let $f(\mathbf{z})$ be a complex function defined in a neighborhood of a point $\mathbf{z}_0$. The *derivative* of $f(\mathbf{z})$ at $\mathbf{z}_0$, denoted by $f'(\mathbf{z}_0)$, is defined by

$$f'(\mathbf{z}_0) = \lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{z}_0 + \mathbf{h}) - f(\mathbf{z}_0)}{\mathbf{h}}$$

assuming that the limit exists.

**Definition.** A function $f(\mathbf{z})$ defined on an open set $\mathbf{U}$ is said to be *analytic* or *holomorphic* on $\mathbf{U}$ if its derivative $f'(\mathbf{z})$ exists at every point of $\mathbf{U}$. A function $f(\mathbf{z})$ defined on an arbitrary set $\mathbf{A}$ is said to be *analytic* or *holomorphic* on $\mathbf{A}$ if it is analytic on an open set containing $\mathbf{A}$.

**E.2.1. Example.** Let n be a positive integer. The function

$$f(\mathbf{z}) = \mathbf{z}^n$$

is analytic on $\mathbf{C}$ and

$$f'(\mathbf{z}) = n\mathbf{z}^{n-1}.$$

Although the definition for the complex derivative looks just like that of the derivative of real functions, it is much more constrained. Let

$$f(\mathbf{z}) = f(x + \mathbf{i}y) = u(x, y) + \mathbf{i}v(x, y). \tag{E.1}$$

If we approach $\mathbf{z}$ first along a line parallel to the x-axis and next by a line parallel to the y-axis, it is easy to show that

$$f'(\mathbf{z}) = \frac{\partial u}{\partial x} + \mathbf{i}\frac{\partial v}{\partial x} = -\mathbf{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}.$$

From this we get the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{E.2}$$

Therefore, that u(x,y) and v(x,y) satisfy the Cauchy-Riemann equations is a necessary condition for $f(\mathbf{z})$ to be differentiable. The converse is essentially also true.

**E.2.2. Theorem.** If the functions u(x,y) and v(x,y) of an arbitrary function $f(\mathbf{z})$ expressed in form (E.1) have continuous partial derivatives in a neighborhood of a point $\mathbf{z}_0$, then $f(\mathbf{z})$ has a derivative at $\mathbf{z}_0$ if and only if the Cauchy-Riemann equations (E.2) hold at $\mathbf{z}_0$.

**Proof.** See [Ahlf66].

One important property of analytic functions is that they preserve the angles and their sense (or orientation or sign) between intersecting curves. More precisely, let $f(\mathbf{z})$ be a function that is analytic at a point $\mathbf{z}_0$. Let $\alpha_1(t)$ and $\alpha_2(t)$ be two curves in the complex plane with $\alpha_1(0) = \alpha_2(0) = \mathbf{z}_0$. Assume that both curves have nonzero tangent vectors at $\mathbf{z}_0$. Define

$$\beta_1(t) = f(\alpha_1(t)),\ \beta_2(t) = f(\alpha_2(t)),\ \text{and}\ \ \mathbf{w}_0 = \beta_1(0) = \beta_2(0).$$

If $f'(\mathbf{z}_0) \neq \mathbf{0}$, then one can show the following:

(1) The angle between the tangent vectors $\alpha_1'(0)$ and $\alpha_2'(0)$ at $\mathbf{z}_0$ is the same as the angle between the tangent vectors $\beta_1'(0)$ and $\beta_2'(0)$ at $\mathbf{w}_0$.
(2) The *sense* (or orientation or sign) of the angles is preserved, that is, if the curves $\alpha_1(t)$ and $\alpha_2(t)$ are not tangent at $\mathbf{z}_0$, then, thinking of the complex plane as $\mathbf{R}^2$, the ordered bases $(\alpha_1'(0),\alpha_2'(0))$ and $(\beta_1'(0),\beta_2'(0))$ induce the same orientation on $\mathbf{R}^2$.

See Figure E.1.

**Definition.** A function $f(\mathbf{z})$ satisfying (1) and (2) at a point $\mathbf{z}_0$ is said to be *conformal at* $z_0$. A *conformal map* is a map that is conformal at every point of its domain.

**E.2.3. Theorem.** An analytic function $f(\mathbf{z})$ is conformal at every point $\mathbf{z}_0$ where $f'(\mathbf{z}_0) \neq \mathbf{0}$.

**Proof.** See [Ahlf66].

In fact, one can show a converse:

**E.2.4. Theorem.** If a function $f(\mathbf{z})$ is conformal on an open set, then it is analytic on that set.

**Proof.** See [Need98].

The main result about real power series carry over to complex ones. A series

$$f(\mathbf{z}) = \sum_{n=0}^{\infty} \mathbf{a}_n \mathbf{z}^n \tag{E.3}$$

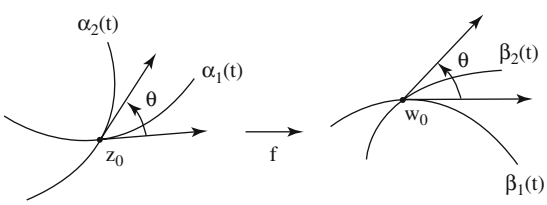



**Figure E.1.** A conformal map.

has a radius of convergence R, $0 \le R \le \infty$ and a disk of convergence. The proofs are the same. An important fact is

**E.2.5. Theorem.** If R is the radius of convergence of the series (E.3), then $f(\mathbf{z})$ is an analytic function for all $\mathbf{z}$, $|\mathbf{z}| < R$. The derivative of $f(\mathbf{z})$ can be obtained by term-wise differentiation and the resulting series has the same radius of convergence.

**Proof.** See [Ahlf66].

**E.2.6. Corollary.** The function $f(\mathbf{z})$ in (E.3) is infinitely differentiable for all $\mathbf{z}$, $|\mathbf{z}| < R$, where R is its radius of convergence.

We can now use series to define some standard functions.

**Definition.** The *exponential function* $e^{\mathbf{z}}$ is defined by

$$e^{\mathbf{z}} = 1 + \frac{\mathbf{z}}{1!} + \frac{\mathbf{z}^2}{2!} + \cdots + \frac{\mathbf{z}^n}{n!} + \cdots.$$

**Definition.** The *sine function* $\sin \mathbf{z}$ and *cosine function* $\cos \mathbf{z}$ are defined by

$$\sin \mathbf{z} = \frac{e^{\mathbf{iz}} + e^{-\mathbf{iz}}}{2} \quad \text{and} \quad \cos \mathbf{z} = \frac{e^{\mathbf{iz}} - e^{-\mathbf{iz}}}{2\mathbf{i}}.$$

It is easy to show that the complex sine and cosine functions have series representations like their real cousins:

$$\sin \mathbf{z} = \mathbf{z} - \frac{\mathbf{z}^3}{3!} + \frac{\mathbf{z}^5}{5!} - \cdots \quad \text{and} \quad \cos \mathbf{z} = 1 - \frac{\mathbf{z}^2}{2!} + \frac{\mathbf{z}^4}{4!} - \cdots$$

**Definition.** A function f(t) defined on **R** or **C** is said to be *periodic of period T* if $f(t + T) = f(t)$.

One can check that the sine and cosine functions are periodic of period $2\pi$. Finally, the definitions also give the famous *Euler formula*

$$e^{\mathbf{i}\theta} = \cos\theta + \mathbf{i}\sin\theta.$$

**Definition.** Any root of the equation $\mathbf{z}^n = 1$ is called an *nth root of unity*.

It should be clear from the above that there are n nth roots of unity. In fact, they are

$$1, \boldsymbol{\omega}, \boldsymbol{\omega}^2, \ldots, \boldsymbol{\omega}^{n-1},$$

where

$$\boldsymbol{\omega} = e^{(2\pi/n)\mathbf{i}} = \cos\frac{2\pi}{n} + \mathbf{i}\sin\frac{2\pi}{n}.$$

## E.3 Complex Integration

The easiest way to define integration of complex-valued functions is to reduce the problem to integration of real-valued functions.

Let $f:[a,b] \to \mathbf{C}$ be a function. Writing

$$f(t) = u(t) + \mathbf{i}v(t),$$

define

$$\int_a^b f(t)dt = \int_a^b u(t)dt + \mathbf{i}\int_a^b v(t)dt.$$

More generally, let $\gamma:[a,b] \to \mathbf{C}$ ($=\mathbf{R}^2$) be a $C^1$ parametric curve and let $f(\mathbf{z})$ be a complex function that is continuous on $\boldsymbol{\Gamma} = \gamma([a,b])$.

**Definition.** Define the *line integral of f along* $\gamma$, $\int_\gamma f$, by

$$\int_\gamma f = \int_\gamma f(\mathbf{z})d\mathbf{z} = \int_a^b f(\gamma(t))\gamma'(t)dt.$$

Using the change of variables formula for integrals it is easy to show that the line integral is invariant under changes of parameter. What this means is the following:

**E.3.1. Theorem.** Let $\phi:[c,d] \to [a,b]$ be a one-to-one $C^1$ map with $\phi' > 0$. If $\lambda:[c,d] \to \boldsymbol{\Gamma}$ is the reparameterization of $\gamma(t)$ defined by $\lambda(t) = \gamma(\phi(t))$, then

$$\int_\lambda f = \int_\gamma f.$$

**Proof.** See [Ahlf66].

A fundamental theorem of complex analysis is

**E.3.2. Theorem.** (The Cauchy Integral Formula) Let $f(\mathbf{z})$ be a function analytic on an open disk $\mathbf{D}$. Let $\gamma:[a,b] \to \mathbf{D}$ be a closed proper $C^1$ parametric curve and let $\mathbf{z}$ be a point of $\mathbf{D}$ not in the image of $\gamma$. Then

$$f(\mathbf{z}) = \frac{1}{2\pi\mathbf{i}}\int_\gamma \frac{f(\zeta)}{\zeta - \mathbf{z}}d\zeta.$$

**Proof.** See [Ahlf66].

**E.3.3. Theorem.** With the same hypotheses as in Theorem E.3.2 one has

$$f^{(n)}(\mathbf{z}) = \frac{1}{2\pi \mathbf{i}} \int_\gamma \frac{f(\zeta)}{(\zeta - \mathbf{z})^{n+1}} d\zeta.$$

**Proof.** One way to prove this theorem is to use Theorem E.3.2 and show that differentiation commutes with integration.

**E.3.4. Corollary.** An analytic function is infinitely differentiable.

## E.4 More on Complex Series

The first theorem is the complex version of the Taylor series expansion for a function.

**E.4.1. Theorem.** If $f(\mathbf{z})$ is an analytic function on an open disk of radius r about a point $\mathbf{z}_0$, then for every point $\mathbf{z}$ in the disk

$$f(\mathbf{z}) = f(\mathbf{z}_0) + f'(\mathbf{z}_0)(\mathbf{z} - \mathbf{z}_0) + \frac{f''(\mathbf{z}_0)}{2!}(\mathbf{z} - \mathbf{z}_0)^2 + \cdots + \frac{f^{(n)}(\mathbf{z}_0)}{n!}(\mathbf{z} - \mathbf{z}_0)^n + \cdots. \tag{E.4}$$

In particular, the series on the right side of equation (E.4) converges.

**Proof.** See [Ahlf66].

Let $f(\mathbf{z})$ be a function that is analytic in the neighborhood of a point $\mathbf{a}$ except possibly at the point $\mathbf{a}$ itself. More precisely, assume that there exists a $\delta > 0$ and that $f(\mathbf{z})$ is analytic for all $\mathbf{z}$ satisfying $0 < |\mathbf{z} - \mathbf{a}| < \delta$.

**Definition.** The point $\mathbf{a}$ is called an *isolated singularity* of $f(\mathbf{z})$. It is called a *removable singularity* of f if some definition of $f(\mathbf{a})$ will make f analytic at $\mathbf{a}$. The point $\mathbf{a}$ is called a *pole* if

$$\lim_{\mathbf{z} \to \mathbf{a}} f(\mathbf{z}) = \infty,$$

and in that case we set $f(\mathbf{a}) = \infty$.

Assume that $f(\mathbf{z})$ has a pole at a point $\mathbf{a}$. If

$$g(\mathbf{z}) = \frac{1}{f(\mathbf{z})},$$

then $g(\mathbf{z})$ has a removable singularity at $\mathbf{a}$ and we can remove the singularity by defining $g(\mathbf{a})$ to be $\mathbf{0}$, making g into a function which is analytic in a neighborhood of $\mathbf{a}$. Because $\mathbf{a}$ is a zero of g, it follows that

$$g(\mathbf{z}) = (\mathbf{z} - \mathbf{a})^m h(\mathbf{z}),$$

for some $m > 0$ and $h(\mathbf{z})$ an analytic function at $\mathbf{a}$ with $h(\mathbf{a}) \neq \mathbf{0}$.

**Definition.** The integer m is called the *order* of the pole of f at **a**.

Since

$$f(\mathbf{z}) = \frac{1}{g(\mathbf{z})},$$

it follows that

$$f(\mathbf{z}) = (\mathbf{z} - \mathbf{a})^{-m} h(\mathbf{z}),$$

where $h(\mathbf{z})$ is analytic in a neighborhood of **a** and $h(\mathbf{a}) \neq \mathbf{0}$.

**Definition.** A *rational* function is a quotient of two polynomials that have no common root.

**Definition.** A *meromorphic* function is a function that is analytic on an open set except possibly for poles.

**E.4.2. Example.** Rational functions are a special case of meromorphic functions.

**E.4.3. Example.** If $f(\mathbf{z})$ and $g(\mathbf{z})$ are analytic functions and if $g(\mathbf{z})$ is not identically **0**, then $f(\mathbf{z})/g(\mathbf{z})$ is a meromorphic function.

Basic facts about series can be applied to complex series with negative powers of **z** such as

$$\sum_{n=0}^{\infty} \mathbf{a}_n \mathbf{z}^{-n}. \tag{E.5}$$

By replacing **z** with $1/\mathbf{z}$ one concludes that the series (E.5) converges for all **z** with $\mathbf{z} > |R|$ for some R and is an analytic function there. Therefore, we can talk about series of the form

$$\sum_{n=-\infty}^{\infty} \mathbf{a}_n \mathbf{z}^{n}, \tag{E.6}$$

where we say that it is convergent if the two parts, the part with positive powers of **z** and the part with negative powers of **z**, are separately convergent. More generally, we can consider series of the form (E.6) where we expand about some arbitrary point $\mathbf{z}_0$. Such series, if they converge, will then converge in some annulus about $\mathbf{z}_0$ and define an analytic function there. The following is a converse to this.

**E.4.4. Theorem.** Let $0 \le R_1 < R_2$ and assume that $f(\mathbf{z})$ is analytic for all **z** in the annulus $R_1 < |\mathbf{z} - \mathbf{z}_0| < R_2$. Then $f(\mathbf{z})$ has a unique representation in that region of the form

$$f(\mathbf{z}) = \sum_{n=1}^{\infty} \frac{\mathbf{a}_n}{(\mathbf{z} - \mathbf{z}_0)^n} + \sum_{n=0}^{\infty} \mathbf{b}_n (\mathbf{z} - \mathbf{z}_0)^n, \tag{E.7}$$

where

$$\mathbf{a}_n = \frac{1}{2\pi \mathbf{i}} \int_{\gamma_1} \frac{f(\mathbf{z})}{(\mathbf{z}-\mathbf{z}_0)^{-n+1}} d\mathbf{z}, \quad n = 1, 2, \ldots,$$

$$\mathbf{b}_n = \frac{1}{2\pi \mathbf{i}} \int_{\gamma_2} \frac{f(\mathbf{z})}{(\mathbf{z}-\mathbf{z}_0)^{n+1}} d\mathbf{z}, \quad n = 0, 1, 2, \ldots,$$

and $\gamma_1$ and $\gamma_2$ are proper parameterizations of circles $\mathbf{C}_1$ and $\mathbf{C}_2$ of radius $r_1$ and $r_2$ about $\mathbf{z}_0$, respectively. The only requirements that the circles have to satisfy are that

$$R_1 < r_2 < |\mathbf{z} - \mathbf{z}_0| < r_1 < R_2$$

and that they are parameterized in a counter-clockwise fashion.

**Proof.** See Figure E.2. This is a fairly straightforward application of Cauchy's integral formula.

**Definition.** The series expansion (E.7) for $f(\mathbf{z})$ is called the *Laurent series* for $f(\mathbf{z})$.

**E.4.5. Example.** The Laurent series expansion for $f(\mathbf{z}) = \dfrac{e^{\mathbf{z}}}{\mathbf{z}^2}$ is

$$\frac{e^{\mathbf{z}}}{\mathbf{z}^2} = \frac{1}{\mathbf{z}^2} + \frac{1}{\mathbf{z}} + \frac{1}{2!} - \frac{\mathbf{z}}{3!} - \frac{\mathbf{z}^2}{4!} + \cdots.$$

## E.5 Miscellaneous Facts

**E.5.1. Theorem.** (Liouville's Theorem) A bounded function that is analytic on the whole plane must be a constant function.

**Proof.** See [Ahlf66].

A fundamental result that is an amazingly trivial consequence of Liouville's theorem is the algebraic closure of the complex numbers:

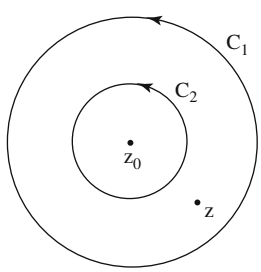



**Figure E.2.** The circles in the Laurent series formulas in Theorem E.4.4.

**E.5.2. Theorem.** (The Fundamental Theorem of Algebra) Every complex polynomial $p(\mathbf{z})$ of degree one or greater has a root.

**Proof.** If $p(\mathbf{z})$ has no zero, then the function $f(\mathbf{z}) = 1/p(\mathbf{z})$ is an analytic function on the whole complex plane. Since it is easy to show that $f(\mathbf{z})$ approaches $\mathbf{0}$ as $|\mathbf{z}|$ goes to infinity, it follows that $f(\mathbf{z})$ is bounded and must therefore be constant by Liouville's theorem. This is impossible and so $p(\mathbf{z})$ must have a zero.

**E.5.3. Corollary.** Every complex polynomial of degree n, $n \geq 1$, factors into n linear factors and has n roots counted with their multiplicity.

**E.5.4. Corollary.** Every real polynomial of degree n, $n \geq 1$, has at most n roots counted with their multiplicity.

Liouville's theorem also implies

**E.5.5. Theorem.** Every function meromorphic on the extended complex plane is a rational function.

**Proof.** See [SakZ71].

**E.5.6. Theorem.** (The Maximum Principle) If $f(\mathbf{z})$ is an analytic function on a closed and bounded set, then the maximum of $|f(\mathbf{z})|$ occurs on the boundary of the set.

**Proof.** See [Ahlf66].

**E.5.7. Theorem.** (Fundamental Theorem of Conformal Mappings) Every simply connected Riemann surface can be mapped in a biholomorphic way (the map and its inverse are holomorphic) onto either the closed plane, the plane, or the interior of the unit disk.

**Proof.** See [BehS62].

**E.5.8. Corollary.** (The Riemann Mapping Theorem) Let $\mathbf{U}$ be a simply connected open subset of the complex plane that is not the entire plane and let $\mathbf{z}_0 \in \mathbf{U}$. Then there exists a unique analytic function $f(\mathbf{z})$ on $\mathbf{U}$ that satisfies $f(\mathbf{z}_0) = \mathbf{0}$, $f'(\mathbf{z}_0) > 0$, and that maps $\mathbf{U}$ in a one-to-one fashion onto the open unit disk.

**Proof.** See [Ahlf66].

APPENDIX F

# A Bit of Numerical Analysis

## F.1 The Condition Number of a Matrix

**F.1.1. Lemma.** If A is an $m \times n$ matrix, then there is a real number $K > 0$, so that

$$|\mathbf{x}A| \le K|\mathbf{x}|$$

for all $\mathbf{x} \in \mathbf{R}^m$.

**Proof.** Assume $A = (a_{ij})$. Let $\mathbf{c}_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$ be the jth column vector and let

$$M = \max\{|\mathbf{c}_1|, |\mathbf{c}_2|, \ldots, |\mathbf{c}_n|\}.$$

If $\mathbf{y} = \mathbf{x}A$, then the Cauchy-Schwarz inequality implies that

$$|y_i| = \left|\sum_{i=1}^{m} a_{ij}x_j\right| \le |\mathbf{c}_i|\,|\mathbf{x}| \le M|\mathbf{x}|,$$

so that

$$|\mathbf{x}A| = |\mathbf{y}| \le \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} \le \sqrt{nM^2|\mathbf{x}|^2} \le \sqrt{n}\mathbf{M}|\mathbf{x}|.$$

We can let $K = \sqrt{n}M$.

**Definition.** Let A be an $m \times n$ matrix. The *norm* of A, denoted by $\|A\|$, is defined by

$$\|A\| = \sup_{\mathbf{0} \ne \mathbf{x} \in \mathbf{R}^n} \frac{|\mathbf{x}A|}{|\mathbf{x}|}.$$

Lemma F.1.1 clearly implies that the norm of an $m \times n$ matrix is well defined because a bounded set of real numbers always has a least upper bound. The next theorem shows that the norm of a matrix behaves like a norm.

**F.1.2. Theorem.** Let A and B be $m \times n$ matrices and let $a \in \mathbf{R}$.

(1) $\|A\| \geq 0$ and $\|A\| = 0$ if and only if A is the zero matrix.
(2) $\|aA\| = |a|\,\|A\|$.
(3) $\|A + B\| \leq \|A\| + \|B\|$.
(4) If $m = n$, then $\|AB\| \leq \|A\|\,\|B\|$.

**Proof.** Straightforward.

**Definition.** Let A be a nonsingular $n \times n$ matrix. The product $\|A\|\,\|A^{-1}\|$ is called the *condition number* of A and is denoted by cond(A).

The condition number of a matrix plays an important role in numerical analysis because it has a direct bearing on the accuracy of numerical solutions to linear systems of the form $\mathbf{x}A = \mathbf{b}$. In this context, one usually says that the linear system or matrix is *ill-conditioned* if cond(A) is "large." Numerical solutions to ill-conditioned systems are typically very inaccurate.

## F.2 Approximation and Numerical Integration

The problem addressed in this section is how, given a function $f:[a,b] \to \mathbf{R}$, one can best approximate the integral

$$I = \int_a^b f \tag{F.1}$$

numerically if no antiderivative of f is available. In no way does this section intend to cover the subject of numerical integration. We simply want to explain the gist of one important technique called Gaussian quadrature because it does come up in geometric modeling. For example, see Chapter 14 of [AgoM05]. For more details see [ConD72] or [Hild87].

We assume that the reader is familiar with the trapezoidal and Simpson's rule from calculus. These methods approximate the integral I in (F.1) by approximating the function f by a straight line and parabola, respectively.

**Trapezoidal rule:** $I \approx \frac{1}{2}(b-a)(f(a)+f(b))$

with the error $E = -\frac{1}{12}f''(\xi)(b-a)^3$ for some $\xi \in (a,b)$

**Simpson's rule:** $I \approx \frac{1}{6}(b-a)\left(f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right)$

with the error $E = -\frac{1}{90}f^{(4)}(\xi)\left(\frac{b-a}{2}\right)^5$ for some $\xi \in (a,b)$

These approximations are usually not very good because the line or parabola are poor approximations to f. The key to improving the accuracy is to look for better approximating polynomial functions. The reason for sticking with polynomials or piecewise polynomial functions is that they are easy to integrate. This leads us of course to another big subject, namely, finding the best piecewise polynomial approximations to a function.

Assume that the function f(x) is known at points $x_0, x_1, \ldots, x_n$. If one looks over some of the simple standard approaches to finding an approximating polynomial, such as Lagrange or Hermite interpolation, one will see that they all try to approximate f(x) by a polynomial g(x) of the form

$$g(x) = \sum_{i=0}^{n} f(x_i)p_i(x), \tag{F.2}$$

where $p_i(x)$ are suitably chosen polynomials. (In the Hermite case one actually adds another similar sum but one using the values of the derivative of f at $x_i$.) Since we want an approximation to f(x), we want the error function

$$E(x) = |f(x) - g(x)| \tag{F.3}$$

to be "small" over [a,b]. Now, we can think of formula (F.2) as representing a linear combination of polynomials $p_i(x)$. Continuing this line of thought, the problem then becomes one of finding a basis $p_0(x), p_1(x), \ldots$ for the subspace of polynomials in the vector space $C^r([a,b])$ so that the linear combination in (F.2) best approximates this arbitrary function f(x). Orthonormal bases of vector spaces always have many advantages and so one is lead to looking for sequences $p_0(x), p_1(x), \ldots$ of "orthogonal" polynomials.

**Definition.** Given an inner product <,> on $C^r([a,b])$, a sequence of polynomials $p_0(x)$, $p_1(x), \ldots$ in $C^r([a,b])$ is called a *sequence of orthogonal polynomials over [a,b]* with respect to <,> if the polynomials $p_i(x)$ have degree i and are pairwise orthogonal with respect to <,>, that is, $\langle p_i(x), p_j(x)\rangle = 0$, for $i \neq j$.

**F.2.1. Theorem.** Let <,> be an inner product on $C^r([a,b])$.

(1) Any sequence of orthogonal polynomials [a,b] with respect to <,> forms a basis for the space of all polynomials over [a,b].
(2) The condition $p_0(x) = 1$ and $p_i(1) = 1$, $i \geq 1$, define a unique sequence of orthogonal polynomials $p_i(x)$ called the *sequence of orthogonal polynomials associated to* <,>. (We assume that 1 belongs to [a,b] here.)

**Proof.** Easy.

The inner product on $C^r([a,b])$ that we have in mind here is

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx, \tag{F.4a}$$

or, more generally,

$$<f,g> = \int_a^b f(x)g(x)w(x)dx, \tag{F.4b}$$

for some nonnegative *weight function* w(x). To keep formulas simple, let us assume that [a,b] = [–1,1]. (It is easy to change the domain of functions by a linear change of variables.)

Note that the sequence 1, x, $x^2$, . . . is not a sequence of orthogonal polynomials over [–1,1] with respect to the inner product defined by (F.4a).

**F.2.2. Theorem.** The polynomials $P_i(x)$ in the sequence of orthogonal polynomials associated to the inner product defined by (F.4a) are defined by the following recursion formulas:

$$\begin{aligned} P_0(x) &= 1 \\ P_{i+1}(x) &= \frac{2i+1}{i+1}xP_i(x) - \frac{i}{i+1}P_{i-1}(x), \quad i \geq 0. \end{aligned} \tag{F.5}$$

**Proof.** See [ConD72].

**Definition.** The orthogonal polynomials $P_i(x)$ defined by formulas (F.5) are called the *Legendre polynomials*.

The first few elements in the sequence of orthogonal polynomials $P_i(x)$ are easily seen to be

$$1,\ x,\ \frac{1}{2}(3x^2-1),\ \frac{1}{2}(5x^3-3x),\ \ldots.$$

One could use a simple Gram-Schmidt type algorithm applied to the sequence 1, x, $x^2$, . . . to find these polynomials.

**F.2.3. Theorem.** The polynomials $T_i(x)$ in the sequence of orthogonal polynomials associated to the inner product defined by (F.4b) with

$$w(x) = \frac{1}{\sqrt{1-x^2}}$$

are defined by

$$T_i(x) = \cos\left(i\cos^{-1}(x)\right) \tag{F.6a}$$

or by the following recursion formulas:

$$\begin{aligned} T_0(x) &= 1 \\ T_{i+1}(x) &= 2xT_i(x) - T_{i-1}(x), \quad i \geq 0. \end{aligned} \tag{F.6b}$$

**Proof.** See [ConD72].

**Definition.** The orthogonal polynomials $T_i(x)$ defined by formulas (F.6) are called the *Chebyshev polynomials*.

The first few Chebyshev polynomials are

$$1,\ x,\ 2x^2-1,\ 4x^3-3x,\ 8x^4-8x^2+1,\ \ldots.$$

One can show that these polynomials give especially good approximations. Particularly relevant for minimizing the error function E(x) in (F.3) is the following fact:

**F.2.4. Theorem.** Of all nth degree polynomials with leading coefficient 1, the polynomials $2^{1-n}T_n(x)$ have the smallest absolute value $2^{1-n}$ on [–1,1].

**Proof.** See [ConD72]. Note that by formula (F.6a) the polynomials $T_i(x)$ are bounded by 1 on [–1,1] since the cosine function is.

Here is how one would use Chebyshev polynomials to get an efficient polynomial approximation to a function f(x) on [–1,1] that is within some ε of the function.

**Step 1:** Begin with the straightforward approach which is to use the Taylor polynomial

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for f(x) of smallest degree n so that the remainder is less than $\varepsilon_n$ for an $\varepsilon_n < \varepsilon$. See Theorem D.2.4.

**Step 2:** Since the monomials $x^i$ are linear combinations of the Chebyshev polynomials $T_i(x)$, replace them by the $T_i(x)$ and express p(x) in the form

$$p(x) = b_0 + b_1T_1(x) + b_2T_2(x) + \cdots + b_nT_n(x).$$

**Step 3:** Choose k to be the smallest integer so that

$$\varepsilon_n + b_{k+1} + \cdots + b_n \le \varepsilon.$$

Then

$$q(x) = b_0 + b_1T_1(x) + b_2T_2(x) + \cdots + b_kT_k(x)$$

will approximate to within ε.

It turns out that this use of Chebyshev polynomials will often produce approximations of degree k of much smaller degree than the degree n of the Taylor polynomial.

We return to the problem of this section. Using the trapezoidal and Simpson's rule as an approximation to the integral I in (F.1) we will get the correct answer if f(x) is a linear or quadratic polynomial respectively. This fact is easy to establish by looking

at the error term E which contains the first, respectively, second derivative as a factor. Thus we are led to ask if we can find an approximation to the integral that gives the exact answer on polynomials of degree $\leq k$. We will achieve this if we can find an approximation whose error term has the $(k + 1)$st derivative as a factor.

In general, given a function $f(x)$ and a weight function $w(x)$, Gaussian quadrature gives an approximation for the integral

$$I = \int_a^b f(x)w(x)dx. \tag{F.7}$$

One uses the inner product on $C^r([a,b])$ defined by (F.4b). Thus there are different flavors of Gaussian quadrature depending on the choice of $w(x)$ in this general setting. *Legendre-Gauss quadrature* (or simply *Gauss* or *Gaussian quadrature* if one is only contemplating the integral (F.1)) computes an approximation of this integral (F.7) when $w(x) = 1$ and uses the inner product defined by (F.4a). *Chebyshev-Gauss quadrature* uses

$$w(x) = \frac{1}{\sqrt{1 - x^2}}.$$

The more general type of Gauss quadrature is useful in situations where one is still trying to approximate an integral of the type (F.1) but the function $f(x)$ has singularities which can be removed by rewriting the integral in the form

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{w(x)} w(x)dx,$$

where

$$F(x) = \frac{f(x)}{w(x)}$$

has no singularities.

**F.2.5. Theorem.** Let $p_i(x)$ be the sequence of orthogonal polynomials with respect to the inner product (F.4b). If $x_0, x_1, \ldots, x_k$ are the zeros of the polynomial $p_{k+1}(x)$ and if we define

$$L_{i,k}(x) = \prod_{j=0, j\neq i}^{k} \frac{x - x_j}{x_i - x_j} \tag{F.8}$$

and

$$a_i = \int_a^b L_{i,k}(x)w(x)dx, \tag{F.9}$$

then the approximation

$$I_k = a_0 f(x_0) + a_1 f(x_1) + \cdots + a_k f(x_k)$$

for the integral I in (F.7) will be exact whenever f(x) is a polynomial function of degree $\leq 2k + 1$. Furthermore,

$$I - I_k = \frac{1}{(2k+2)!} f^{(2k+2)}(\xi) \frac{R_{k+1}}{c_{k+1}^2}$$

for some $\xi \in (a,b)$, where

$$R_{k+1} = \int_a^b (p_{k+1}(x))^2 w(x) du$$

and $c_{k+1}$ is the coefficient of the leading term in $p_{k+1}(x)$.

**Proof.** See [ConD72].

Theorem F.2.5 is one of the basic Gaussian quadrature results. The polynomials $L_{i,k}(x)$ in formula (F.8) are just the *Lagrange basis functions* that are used in Lagrange interpolation.

**Definition.** The zeros of the polynomials in the sequence of orthogonal polynomials with respect to the inner product (F.4b) are called the *Gaussian points* or *zeros*. The corresponding coefficients $a_i$ defined by equation (F.9) are called the *Gaussian weights*.

Now, since the sequences of orthogonal polynomials are well known, so are their Gaussian zeros and weights, which can be looked up in a table. Therefore, to use Gaussian quadrature one only has to be able to evaluate the function f(x) at the Gaussian zeros. For example, in the case of Legendre-Gauss quadrature with k = 1, the zeros of

$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$

are

$$x_0 = -\frac{1}{\sqrt{3}} \quad \text{and} \quad x_1 = \frac{1}{\sqrt{3}}.$$

Also,

$$a_0 = \int_{-1}^{1} \frac{x - \frac{1}{\sqrt{3}}}{\left(-\frac{1}{\sqrt{3}}\right) - \frac{1}{\sqrt{3}}} dx = 1 \qquad \text{and} \qquad a_1 = \int_{-1}^{1} \frac{x - \left(-\frac{1}{\sqrt{3}}\right)}{\frac{1}{\sqrt{3}} - \left(-\frac{1}{\sqrt{3}}\right)} dx = 1.$$

Therefore, the two-point Gaussian quadrature approximation is

$$\int_{-1}^{1} f \approx a_0 f(x_0) + a_1 f(x_1) = f\left(-\frac{1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right).$$

One can also show that the error in our approximation is

$$\frac{1}{135} f^{(4)}(\xi)$$

for some $\xi \in (-1,1)$.

In summary, Gaussian quadrature gives extremely good accuracy for smooth functions with a relative few number of function evaluations. It often also works well for functions that are not so well behaved. It cannot in general be used if only a predetermined finite number of values of the function are known because one needs the values at the Gaussian zeros.

# Bibliography

## Abbreviations

| | |
|---|---|
| ACM TOG | ACM Transactions on Graphics |
| AMS | American Mathematical Society |
| CAD | Computer Aided Design |
| CAGD | Computer Aided Geometric Design |
| CG&A | IEEE Computer Graphics & Applications |
| SIGGRAPH | Proceedings of yearly SIGGRAPH conference which is the July or August issue of Computer Graphics |

## Abstract Algebra


[Dean66] Dean, Richard A., *Elements of Abstract Algebra*, John Wiley & Sons, Inc., 1966.

[Fral67] Fraleigh, John B., *A First Course in Abstract Algebra*, Addison-Wesley Publ. Co., 1967.

[Jaco66] Jacobson, Nathan, *Lectures in Abstract Algebra*, Volume I – Basic Concepts, D. Van Nostrand Co., Inc., 1966.

[Jaco64] Jacobson, Nathan, *Lectures in Abstract Algebra*, Volume III – Theory of Fields and Galois Theory, D. Van Nostrand Co., Inc., 1964.

[Mill58] Miller, Kenneth S., *Elements of Modern Abstract Algebra*, Harper & Brothers, Publishers, 1958.

[Waer50] Van der Waerden, B.L., *Modern Algebra*, Volume I, Frederick Ungar Publ. Co., 1950.


## Advanced Calculus


[Buck78] Buck, R.C., *Advanced Calculus*, 3rd Edition, McGraw-Hill Book Co., 1978.

[CodL55] Coddington, Earl A., and Levinson, Norman, *Theory of Ordinary Differential Equations*, McGraw-Hill Book Co., Inc., 1955.

[Flan63] Flanders, Harley, *Differential Forms with Applications to the Physical Sciences*, Academic Press, 1963.

[Spiv65] Spivak, M., *Calculus on Manifolds*, W.A. Benjamin, Inc., 1965.

[TayM72] Taylor, A.E., and Mann, W.R., *Advanced Calculus*, 2nd Edition, John Wiley & Sons, Inc., 1972.


## Algebraic Curves and Surfaces


[Baja92a] Bajaj, Chanderjit, "Algebraic Surface Design and Finite Element Meshes," Technical Report CSD-TR-92-011, Comp. Science Dept., Purdue Univ., West Lafayette, Indiana, 47907-1398, USA, February 1992.

[Baja92b] Bajaj, Chanderjit, "The Emergence of Algebraic Curves and Surfaces in Geometric Design," Technical Report CSD-TR-92-052, Comp. Science Dept., Purdue Univ., West Lafayette, Indiana, 47907-1398, USA, September 1992.

[Hobb91] Hobby, John D., "Numerically Stable Implicitation of Cubic Curves," ACM TOG, **10**(3), July 1991, 255–296.
[Hoff93] Hoffmann, Christoph M., "Implicit Curves and Surfaces in CAGD," CG&A, **13**(1), January 1993, 79–88.
[ManC92b] Manocha, Dinesh, and Canny, John F., "Algorithm for Implicitizing Rational Parametric Surfaces," CAGD, **9**(1), May, 1992, 25–50.

## Algebraic Geometry

[Abhy90] Abhyankar, Shreeram S., *Algebraic Geometry for Scientists and Engineers*, AMS, 1990.
[AdaL94] Adams, William W., and Loustaunau, Philippe, *An Introduction to Gröbner Bases*, AMS, 1994.
[Arno83] Arnon, D.S., "Topologically Reliable Displays of Algebraic Curves," SIGGRAPH 83, **17**(3), July 1983, 219–227.
[BriK81] Brieskorn, E., and Knörrer, H., *Ebene Algebraische Kurven*, Birkhäuser, 1981.
[CoLO97] Cox, David, Little, John, and O'Shea, Donal, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 2nd Edition, Springer-Verlag, 1997.
[CoLO98] Cox, David, Little, John, and O'Shea, Donal, *Using Algebraic Geometry*, Springer-Verlag, 1998.
[Fult69] Fulton, William, *Algebraic Curves*, W.A. Benjamin, Inc., 1969.
[Harr92] Harris, Joe, *Algebraic Geometry: A First Course*, Springer-Verlag, 1992.
[Kend77] Kendig, Keith, *Elementary Algebraic Geometry*, Springer-Verlag, 1977.
[KleL72] Kleinman, S.L., and Laksov, Dan, "Schubert Calculus," Amer. Math. Monthly, **79**(10), December 1972, 1061–1082.
[Mish93] Mishra, Bhubaneswar, *Algorithmic Algebra*, Springer-Verlag, 1993.
[Sede87] Sederberg, Thomas W., "Algebraic Geometry for Surface and Solid Modeling," in [Fari87], 29–42.
[Seid68] Seidenberg, A., *Elements of the Theory of Algebraic Curves*, Addison-Wesley Publ. Co., 1968.
[Shaf94] Shafarevich, Igor R., *Basic Algebraic Geometry I*, 2nd Edition, Springer-Verlag, 1994.
[Walk50] Walker, Robert J., *Algebraic Curves*, Princeton University Press, 1950.
[ZarS60] Zariski, Oscar, and Samuel, Pierre, *Commutative Algebra: Volume I and II*, D. Van Nostrand Co., Inc., 1960.

## Algebraic Topology

(See also Topology)

[AgoM76] Agoston, Max K., *Algebraic Topology: A First Course*, Marcel Dekker, New York, 1976.
[Arms83] Armstrong, M. A., *Basic Topology*, Springer-Verlag, 1983.
[Cair68] Cairns, Stewart S., *Introductory Topology*, The Ronald Press Co., New York, 1968.
[CooF67] Cooke, George E., and Finney, Ross L., *Homology of Cell Complexes*, Princeton Univ. Press, 1966.
[CroF65] Crowell, Richard H., and Fox, Ralph H., *Introduction to Knot Theory*, Blaisdell Publ. Co., 1965.
[Dieu89] Dieudonné, Jean, *A History of Algebraic and Differential Topology 1900–1960*, Birkhäuser, 1989.
[FreF67] Frechet, Maurice, and Fan, Ky, *Initiation to Combinatorial Topology*, Prindle, Weber & Schmidt, 1967.
[Gott96] Gottlieb, Daniel Henry, "All the Way with Gauss-Bonnet and the Sociology of Mathematics," Amer. Math. Monthly, **103**(6), June–July 1996, 457–469.
[GrüS94] Grünbaum, B., and Shephard, G.C., "A New Look at Euler's Theorem for Polyhedra," Amer. Math. Monthly, **2**(101), February 1994, 109–128.
[HilP94] Hilton, P.J., and Pederson, J., "Euler's Theorem for Polyhedra: A Topologist and a Geometer Respond," Amer. Math. Monthly, **101**(10), December 1994, 959–962.
[HilP96] Hilton, P.J., and Pederson, J., "The Euler Characteristic and Pólya's Dream," Amer. Math. Monthly, **103**(2), February 1996, 121–151.
[HilW60] Hilton, P.J., and Wylie, S., *Homology Theory: An Introduction to Algebraic Topology*, Cambridge Univ. Press, 1960.
[Huse66] Husemoller, Dale, *Fibre Bundles*, McGraw-Hill Book Co., 1966.
[Livi93] Livingston, Charles, *Knot Theory*, Math. Association of America, 1993.
[Mark58] Markov, A.A., "Unsolvability of the Problem of Homeomorphy" (Russian), Proceedings of International Congress of Mathematicians, 1958, 300–306.

[Mass67] Massey, W.S., *Algebraic Topology: An Introduction*, Harcourt, Brace, Jovanovich, 1967.
[Matv03] Matveev, S., *Algorithmic Topology and Classification of 3-Manifolds*, Springer-Verlag, 2003.
[McCl85] McCleary, John, *User's Guide to Spectral Sequences*, Publish or Perish, Inc., 1985.
[MilS74] Milnor, John, and Stasheff, Jamed D., *Characteristic Classes*, Princeton Univ. Press and University of Tokyo Press, 1974.
[Mois77] Moise, Edwin E., *Geometric Topology in Dimensions 2 and 3*, Springer-Verlag, 1977.
[Reid35] Reidemeister, K., "Homotopieringe und Linsenräume," Abh. Math. Semin. Univ. Hamburg, **11**, 1935, 102–109.
[Rolf76] Rolfsen, Dale, *Knots and Links*, Publish or Perish, Inc., 1976.
[SeiT80] Seifert, H., and Threlfall, W., *Seifert and Threlfall: A Textbook of Topology*, translated by Michael A. Goldman, Academic Press, 1980.
[Span66] Spanier, Edwin H., *Algebraic Topology*, McGraw-Hill Book Co., 1966.
[Stee51] Steenrod, Norman, *The Topology of Fiber Bundles*, Princeton Univ. Press, 1951.
[Tiet08] Tietze, H., "Über die Topologischen Varianten Mehrdimensionaler Mannigfaltigkeiten," Monatsh. für Math. u. Phys., **19**, 1908, 1–118.
[Whit41] Whitehead, J.H.C., "On Incidence Matrices, Nuclei and Homotopy Types," Ann. of Math., **42**, 1941, 1197–1239.

## Analytic Geometry

[Full73] Fuller, Gordon, *Analytic Geometry*, 4th Edition, Addison-Wesley Publ. Co., 1973.

## Complex Analysis

[Abik81] Abikoff, William, "The Uniformization Theorem," Amer. Math. Monthly, **88**(8), October 1981, 574–592.
[Ahlf66] Ahlfors, Lars V., *Complex Analysis*, 2nd Edition, McGraw-Hill Book Co., 1966.
[BehS62] Behnke, H.C. Heinrich, and Sommer, Friedrich, *Theorie der Analytischen Funktionen einer Komplexen Veränderlichen*, 2nd Edition, Springer-Verlag, 1962.
[Need98] Needham, Tristan, *Visual Complex Analysis*, Oxford University Press, 1998.
[SakZ71] Saks, Stanislaw, and Zygmund, Antoni, *Analytic Functions*, 3rd Edition, Elsevier Publ. Co., 1971.
[Weyl55] Weyl, Hermann, *The Concept of a Riemann Surface*, Addison-Wesley Publ. Co., 1955.

## Conics

[Eise39] Eisenhart, Luther P., *Coordinate Geometry*, Ginn and Company, 1939.

## Cyclides

[Boeh89] Boehm, Wolfgang, "Some Remarks on Cyclides in Solid Modeling," in [Strasser, Wolfgang, and Seidel Hans-Peter, editors, *Theory and Practice of Geometric Modeling*, Springer-Verlag, 1989], 247–252.
[Boeh90] Boehm, Wolfgang, "On Cyclides in Geometric Modeling," CAGD, **7**(1–4), June 1990, 243–255.
[ChDH89] Chandru, V., Dutta, D., and Hoffmann, C.M., "On the Geometry of Dupin Cyclides," The Visual Computer, **5**(5), October 1989, 277–290.
[Dupi22] Dupin, C.P., *Applications de Géométrie et de Méchanique*, Bachelier, Paris, 1822.
[MaPS86] Martin, R.R., de Pont, J., and Sharrock, T.J., "Cyclide Surfaces in Computer Aided Design," in [Greg86], 253–267.
[Mart82] Martin, R.R., "Principal Patches for Computational Geometry," PhD thesis, Engineering Department, Cambridge University, U.K., 1982.
[Maxw68] Maxwell, J.C., "On the Cyclide," Quart. J. Pure Appl. Math., **9**, 1868, 111–126.
[PalB98] Paluszny, Marco, and Boehm, Wolfgang, "General Cyclides," CAGD, **15**(7), July 1998, 699–710.

[Prat90] Pratt, M.J., "Cyclides in Computer Aided Geometric Design," CAGD, **7**(1–4), June 1990, 221–242.

## Differential Geometry

[BaGM82] Banchoff, T., Gaffney, T., and McCrory, C., *Cusps of Gauss Mappings*, Pittman Publ. Inc., 1982.
[Berg03] Berger, Marcel, *A Panoramic View of Riemannian Geometry*, Springer-Verlag, 2003.
[Brec92] Brechner, Eric L., "General Offset Curves and Surfaces," in [Barnhill, Robert E., editor, *Geometry Processing for Design and Manufacturing*, SIAM, 1992], 101–121.
[Call86] Calladine, C.R., "Gaussian Curvature and Shell Structures," in [Greg86], 179–196.
[DoCa76] do Carmo, Manfredo P., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Inc., 1976.
[Gluc71] Gluck, H., "The Converse to the Four Vertex Theorem," L'Enseignement Mathèmatique T.XVII, fasc. 3–4(1971), 295–309.
[Gray98] Gray, Alfred, *Modern Differential Geometry of Curves and Surfaces with MATHEMATICA*, 2nd Edition, CRC Press, 1998.
[Hick65] Hicks, Noel J., *Notes on Differential Geometry*, D. Van Nostrand Co., Inc., 1965.
[KobN63] Kobayashi, Shoshichi, and Nomizu, Katsumi, *Foundations of Differential Geometry*, Volume 1, Interscience Publishers (a division of John Wiley & Sons), 1963.
[Laug65] Laugwitz, D., *Differential and Riemannian Geometry*, Academic Press, 1965.
[Lips69] Lipschutz, M.M., *Differential Geometry*, Schaum's Outline Series, McGraw-Hill, 1969.
[McCl97] McCleary, John, *Geometry from a Differentiable Viewpoint*, Cambridge Univ. Press, 1997.
[MilP77] Millman, Richard S., and Parker, George D., *Elements of Differential Geometry*, Prentice-Hall, 1977.
[ONei66] O'Neill, B., *Elementary Differential Geometry*, Academic Press, 1966.
[Osse69] Osserman, Robert, *A Survey of Minimal Surfaces*, Van Nostrand Reinhold Co., 1969.
[Spiv70a] Spivak, M., *A Comprehensive Introduction to Differential Geometry*, Volume I, Publish or Perish, Inc., 1970.
[Spiv70b] Spivak, M., *A Comprehensive Introduction to Differential Geometry*, Volume II, Publish or Perish, Inc., 1970.
[Spiv75] Spivak, M., *A Comprehensive Introduction to Differential Geometry*, Volume III, Publish or Perish, Inc., 1975.
[Stok69] Stoker, J.J., *Differential Geometry*, Wiley-Interscience, 1969.
[Taba00] Tabachnikov, Serge, "A Four Vertex Theorem for Polygons," Amer. Math. Monthly, **107**(9), November 2000, 830–833.
[Thor79] Thorpe, J.A., *Elementary Topics in Differential Geometry*, Springer-Verlag, 1979.
[Wein00] Weiner, Joel L., "How Helical Can a Closed, Twisted Space Curve Be?," Amer. Math. Monthly, **107**(4), April 2000, 327–333.
[Will59] Willmore, T.J., *An Introduction to Differential Geometry*, Oxford Univ. Press, 1959.

## Differential Topology

[AgoM76b] Agoston, Max K., "Twenty Years of Differential Topology: Some Highlights," Math. Chronicle, **4**, 1976, 77–89.
[AusM63] Auslander, Louis, and MacKenzie, Robert E., *Introduction to Differentiable Manifolds*, McGraw-Hill Book Co., Inc., 1963.
[BisC64] Bishop, Richard L., and Crittenden, Richard J., *Geometry of Manifolds*, Academic Press, 1964.
[FreL89] Freedman, Michael H., and Luo, Feng, *Selected Applications of Geometry to Low-Dimensional Topology*, AMS, 1989.
[GuiP74] Guillemin, Victor, and Pollack, Alan, *Differential Topology*, Prentice-Hall, Inc., 1974.
[Hirs76] Hirsch, Morris W., *Differential Topology*, Springer-Verlag, 1976.
[Miln58] Milnor, John, "Differential Topology," Notes, Princeton Univ., 1958.
[Miln63] Milnor, John, *Morse Theory*, Princeton Univ. Press, 1963.
[Miln65a] Milnor, John, *Lectures on the h-Cobordism Theorem*, Princeton Univ. Press, 1965.
[Miln65b] Milnor, John, *Topology from the Differentiable Viewpoint*, The University Press of Virginia, 1965.
[Miln03] Milnor, John, "Towards the Poincaré Conjecture and the Classification of 3-Manifolds," Notices of the AMS, 50(10), November 2003, 1226–1233.
[Munk61] Munkres, James R., *Elementary Differential Topology*, Princeton Univ. Press, 1961.

[Smal61] Smale, S., "Generalized Poincaré's Conjecture in Dimensions Greater Than Four," Ann. of Math., **75**(2), 1961, 391–406.
[Smal62] Smale, S., "On the Strutcture of Manifolds," Amer. J. Math., **84**, 1962, 387–399.
[Thom98] Thompson, Abigail, "Algorithmic Recognition of 3-Manifolds," Bull. AMS, **35**(1), 1998, 57–66.
[Wall68] Wallace, Andrew H., *Differential Topology: First Steps*, W. A. Benjamin, Inc., 1968.

## Geodesics

[Lyus64] Lyusternik, L.A., *Shortest Paths: Variational Problems*, The MacMillan Co., New York, 1964.

## Geometric Modeling

[AgoM05] Agoston, Max K., *Computer Graphics and Geometric Modeling: Implementation and Algorithms*, Springer, 2005.

## Linear Algebra

[BanW83] Banchoff, Thomas, and Wermer, John, *Linear Algebra Through Geometry*, Springer-Verlag, 1983.
[Brad75] Bradley, Gerald L., *A Primer of Linear Algebra*, Prentice-Hall, 1975.
[Divi75] Divinsky, Nathan, *Linear Algebra*, Page Ficklin Publ. Co., 1975.
[Fink72] Finkbeiner, Daniel T., II, *Elements of Linear Algebra*, W.H. Freeman and Co., 1972.
[HofK71] Hoffman, Kenneth, and Kunze, Ray, *Linear Algebra*, 2nd Edition, Prentice-Hall, Inc., 1971.
[IpsM95] Ipsen, Ilse C.F., and Meyer, Carl D., "The Angle Between Complementary Subspaces," Amer. Math. Monthly, **102**(10), December 1995, 904–911.
[John67] Johnson, Richard E., *Linear Algebra*, Prindle, Weber & Schmidt, Inc., 1967.
[Lips68] Lipschutz, Seymour, *Linear Algebra*, McGraw-Hill Book Co., 1968.
[NobD77] Noble, Ben, and Daniel, James W., *Applied Linear Algebra*, 2nd Edition, Prentice-Hall, Inc., 1977.
[Penr55] Penrose, R., "A Generalized Inverse for Matrices," Proc. Cambridge Phil. Soc., **51**, 1955, 406–413.
[RaoM71] Rao, C.R., and Mitra, S.K., *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, 1971.

## Miscellaneous

[AppH77] Appel, K., and Haken, W., "The Solution of the Four-Color Map Problem," Sci. Amer., **237**, 1977, 108–121.
[BoeP94] Boehm, Wolfgang, and Prautzsch, Hartmut, *Geometric Concepts for Geometric Design*, A K Peters, Ltd., 1994.
[Fari87] Farin, Gerald, editor, *Geometric Modeling: Algorithms and New Trends*, SIAM, 1987.
[Greg86] Gregory, J.A., editor, *The Mathematics of Surfaces*, Clarendon Press, Oxford, 1986.
[HilC99] Hilbert, D., and Cohn-Vossen, S., *Geometry and the Imagination*, 2nd Edition, AMS Chelsea Publ., 1999.
[Hoff89] Hoffmann, Christoph M., *Geometric and Solid Modeling: An Introduction*, Morgan Kaufmann Publ. Inc., 1989.
[RogA90] Rogers, D.F., and Adams, J.A., *Mathematical Elements for Computer Graphics*, 2nd Edition, McGraw-Hill, 1990.

## Numerical Methods

[AbdY96] Abdel-Malek, Karim, and Yeh, Harn-Jou, "Determining Intersection Curves Between Surfaces of Two Solids," CAD, **28**(6/7), June/July 1996, 539–549.
[Alex96] Alexander, Dan, "Newton's Method – Or Is It?," FOCUS, Math. Association of America, October 1996, 32–33.
[ConD72] Conte, S.D., and de Boor, C., *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw-Hill Book Co., 1972.
[DahB74] Dahlquist, G., and Bjorck, A., *Numerical Methods*, Prentice-Hall, Inc., 1974.

[ForM67] Forsythe, George E., and Moler, Cleve B., *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Inc., 1967.
[Hild87] Hildebrand, F.B., *Introduction to Numerical Analysis*, 2nd Edition, Dover Publications, Inc., 1987.

## Offset Curves and Surfaces

[FarN90a] Farouki, R.T., and Neff, C.A., "Analytic Properties of Plane Offset Curves," CAGD, **7**, 1990, 83–99.

## Projective Geometry and Transformations

(motions, affine maps, projectivities)

[Ayre67] Ayres, Frank Jr., *Projective Geometry*, Schaum's Outline Series, McGraw-Hill Book Co., 1967.
[Blin98] Blinn, James F., "The Cross Ratio," CG&A, **18**(6), November/December 1998, 78–80.
[Fari95] Farin, Gerald, *NURB Curves and Surfaces: From Projective Geometry to Practical Use*, A.K. Peters, 1995.
[Gans69] Gans, D., *Transformations and Geometries*, Appleton-Century-Crofts, 1969.
[PenP86] Penna, M.A., and Patterson, R.R., *Projective Geometry and its Applications to Computer Graphics*, Prentice-Hall, 1986.

## Quadrics

[PetR98] Peters, Jörg, and Reif, Ulrich, "The 42 Equivalence Classes of Quadratic Surfaces in Affine n-Space," CAGD, **15**(5), May, 1998, 459–473.

## Real Analysis

[Berb66] Berberian, Sterling K., *Measure and Integration*, The Macmillan Co., 1966.
[Nata61] Natanson, I.P., *Theory of Functions of a Real Variable*, translated from the Russian by Leo F. Boron, Frederick Ungar Publ. Co., 1961.
[Spie69] Spiegel, Murray R., *Theory and Problems of Real Variables*, Schaum's Outline Series, McGraw-Hill, Co., 1969.

## Topology

(See also Algebraic and Differential Topology)

[Barr72] Barr, Stephen, *Experiments in Topology*, Thomas Y. Crowell Co., Apollo Edition, 1972.
[BurM71] Burgesser, H., and Mani, P., "Shellable Decompositions of Cells and Spheres," Math. Scand., **29**, 1971, 197–205.
[ChiS66] Chinn, W. G., and Steenrod, N. E., *First Concepts of Topology*, Random House, 1966.
[Crom97] Cromwell, Peter R., *Polyhedra*, Cambridge Univ. Press, 1997.
[Eise74] Eisenberg, Murray, *Topology*, Holt, Rinehart and Winston, 1974.
[Free82] Freedman, M., "The Topology of Four-Dimensional Manifolds," J. Differential Geom., **17**, 1982, 357–453.
[Jäni84] Jänich, Klaus, *Topology*, Springer-Verlag, 1984.
[KirS69] Kirby, R.C., and Siebenmann, L.C., "On the Triangulation of Manifolds and the Hauptvermutung," Bull. AMS, **75**, 1969, 742–749.
[Lips65] Lipschutz, Seymour, *General Topology*, Schaum's Outline Series, McGraw-Hill, 1965.
[LunW69] Lundell, Albert T., and Weingram, Stephen, *The Topology of CW Complexes*, Van Nostrand Reinhold Co., 1969.
[Mois52] Moise, E.E., "Affine Structures in 3-Manifolds: V," Ann. Math., **56**(2), 1952, 96–114.
[Radó25] Radó, T., "Über des Begriff der Riemannschen Fläche," Acta. Sci. Math. Szeged. **2**, 1925, 101–121.
[Schu68] Schubert, Horst, *Topology*, Allyn and Bacon, Inc., 1968.

# Index

**G**

# Bibliographic Index

[Eise39], 167, 168, 176, 917
[Eise74], 216, 217, 281, 287, 289, 290, 291, 294, 296, 297, 299, 300, 301, 302, 305, 306, 307, 308, 309, 313, 920

[Fari87], 916, 919
[Fari95], 199, 920
[FarN90a], 587, 588, 589, 920
[Fink72], 48, 49, 919
[Flan63], 265, 915
[ForM67], 57, 920
[Fral67], 830, 836, 838, 915
[Free82], 498, 920
[FreF67], 916
[FreL89], 503, 918
[Full73], 168, 917
[Fult69], 808, 916

[Gans69], 66, 81, 99, 101, 126, 134, 135, 136, 152, 199, 920
[Gluc71], 570, 918
[Gott96], 324, 916
[Gray98], 569, 570, 571, 572, 578, 600, 609, 639, 918
[Greg86], 917, 918, 919
[GrüS94], 324, 916
[GuiP74], 265, 536, 548, 918

[Harr92], 775, 776, 789, 793, 796, 804, 916
[Hick65], 633, 918
[HilC99], 166, 167, 168, 192, 200, 201, 202, 203, 204, 205, 557, 600, 604, 640, 642, 919
[Hild87], 908, 920
[HilP94], 324, 916
[HilP96], 324, 916
[HilW60], 399, 433, 916
[Hirs76], 504, 519, 527, 528, 531, 532, 533, 534, 535, 918
[Hobb91], 728, 916
[Hoff89], 789, 790, 919
[Hoff93], 744, 916
[HofK71], 873, 919
[Huse66], 515, 551, 916

[IpsM95], 18, 919

[Jaco64], 856, 857, 858, 915
[Jaco66], 840, 845, 846, 915
[Jäni84], 313, 314, 315, 392, 427, 428, 429, 430, 431, 434, 920
[John67], 854, 865, 867, 868, 869, 919

[Kend77], 678, 683, 688, 689, 764, 791, 792, 793, 795, 798, 799, 800, 801, 802, 803, 916
[KirS69], 339, 920
[KleL72], 796, 916
[KobN63], 875, 880, 918

[Laug65], 649, 918
[Lips65], 281, 288, 920
[Lips68], 14, 44, 108, 861, 865, 866, 919
[Lips69], 576, 577, 584, 585, 598, 618, 620, 649, 670, 918
[Livi93], 420, 916
[LunW69], 391, 392, 438, 920
[Lyus64], 621, 919

[ManC92b], 916
[MaPS86], 917
[Mark58], 453, 916
[Mart82], 917
[Mass67], 354, 417, 418, 420, 428, 429, 430, 431, 433, 500, 917
[Matv03], 452, 917
[Maxw86], 640, 917
[McCl85], 453, 917
[McCl97], 633, 918
[Mill58], 817, 839, 847, 848, 849, 850, 856, 915
[Miln58], 504, 918
[Miln63], 251, 491, 492, 495, 918
[Miln65a], 491, 497, 498, 918
[Miln65b], 532, 535, 918
[Miln03], 499, 918
[MilP77], 609, 610, 632, 663, 918
[MilS74], 449, 487, 535, 551, 917
[Mish93], 790, 916
[Mois52], 339, 920
[Mois77], 452, 917
[Munk61], 231, 468, 504, 551, 918

[Nata61], 896, 920
[Need98], 898, 900, 917
[NobD77], 14, 865, 868, 919

[ONei66], 605, 656, 658, 918
[Osse69], 609, 918

# Index of Algorithms