INTRODUCTION TO LIE GROUPS AND LIE ALGEBRAS

ARTHUR A. SAGLE
RALPH E. WALDE

ACADEMIC PRESS New York and London 1973

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 72-77350

AMS (MOS) 1970 Subject Classifications: 22E15, 22E60, 17B05

PRINTED IN THE UNITED STATES OF AMERICA

Preface

This text is intended for the beginning graduate student with minimal preparation. However since Lie groups abstract the analytic properties of matrix groups, the student is expected to have some knowledge of senior level algebra, topology, and analysis as given in some of the references. In Chapter 1 we review some advanced calculus and extend these results to manifolds in Chapter 2. Consequently the reader knowing these results can skip these chapters but should pay attention to the examples on matrix groups. After this the reader probably should follow the order given in the contents noting that the first part of the text is about Lie groups while the algebraic study of Lie algebras begins in Chapter 9. We have not attempted to prove all basic results so the serious student should take the indicated detours to such texts as those by Freudenthal and de Vries, Helgason, Jacobson, or Wolf. In particular the student must develop his own taste in this subject and ours is only one point of view.

We are deeply indebted to the many authors and teachers in the subject and to them we express our gratitude. We would like to thank Dr. J. R. Schumi for his assistance, Professors Harry Allen, Charles Conatser, and Earl Taft for their timely suggestions, and the National Science Foundation for their support of mathematics. Also thanks are due to Barb Ketter and Els Sagle for their excellent manuscript preparation and to the people at T.S.V. and at the University of Minnesota for their encouragement.

CHAPTER 1

SOME CALCULUS

We shall present many familiar concepts of differential calculus in the terminology of linear algebra. Thus, for functions from one Euclidean space to another, derivatives are given as linear transformations, higher order derivatives are given as multilinear forms, and Taylor's series is presented in this terminology. Instead of giving detailed proofs we present many examples involving matrix groups which will be abstracted in later chapters.

1. Basic Notation

We now informally review some basic concepts with which the reader should be familiar. Thus let V denote an n-dimensional vector space over a field K and let $X_1, \ldots, X_n$ be a basis of V. Then any **point** or **vector** p (or P) in V can be uniquely represented by

$$p = \sum_{i=1}^{n} p_i X_i$$

and we call the $p_i \in K$ the **coordinates** of p relative to the basis $X_1, \ldots, X_n$ of V. In particular when we let $e_k = (0, \ldots, 1, 0 \ldots 0)$ with 1 in the kth position and use the basis $e_1, \ldots, e_n$, then we frequently write $p = (p_1, \ldots, p_n)$ as a vector in V. For a fixed basis $X_1, \ldots, X_n$ of V, the functions

$$u_i : V \to K : p \to p_i$$

for $i = 1, \ldots, n$ are called the **coordinate functions** for V relative to the basis $X_1, \ldots, X_n$. Thus we obtain a "coordinate system" by a choice of basis in V and in particular obtain the usual coordinate system by choosing the $e_1, \ldots, e_n$ basis of V.

Let W be an m-dimensional vector space over K and with V as above let

$$\mathrm{Hom}_K(V, W)$$

or just $\mathrm{Hom}(V, W)$ denote the set of linear transformations of V into W. Thus

$$T : V \to W : X \to T(X)$$

is in $\mathrm{Hom}(V, W)$ if $T(aX + bY) = aT(X) + bT(Y)$ for all $a, b \in K$ and $X, Y \in V$. In particular $\mathrm{Hom}(V, W)$ is a vector space over K of dimension $m \cdot n$ relative to the usual definitions: For $S, T \in \mathrm{Hom}(V, W)$ and $a, b \in K$ define $(aS + bT)(X) = aS(X) + bT(X)$ for all $X \in V$. We shall also use the notation

$$L(V, W) \qquad \text{for} \quad \mathrm{Hom}(V, W)$$

and

$$\mathrm{End}(V) \qquad \text{for} \quad \mathrm{Hom}(V, V).$$

Now let $K = R$, the real numbers, and let $V = R^n$ which we regard as the set of all n-tuples $X = (x_1, \ldots, x_n)$ with $x_i \in R$ and with the operations

$$aX + bY = (ax_1 + by_1, \ldots, ax_n + by_n)$$

for $Y = (y_1, \ldots, y_n)$ and $a, b \in R$. With this representation of V we have a natural **inner product**

$$B : V \times V \to R$$

given by the formula

$$B(X, Y) = \sum_{i=1}^{n} x_i y_i .$$

Thus for $X, Y, Z \in V$ and $a, b \in R$, B satisfies

(1) $B(aX + bY, Z) = aB(X, Z) + bB(Y, Z)$;
(2) $B(X, Y) = B(Y, X)$;
(3) $B(X, X) \geq 0$ and $B(X, X) = 0$ if and only if $X = 0$.

Any function $B : V \times V \to R$ satisfying (1) and (2) above is called a **symmetric bilinear form** and if it also satisfies (3), B is called a **positive definite symmetric bilinear form.**

A **norm** on a vector space V is a function $n : V \to R$ satisfying

(1) $n(X) \geq 0$ and $n(X) = 0$ if and only if $X = 0$;
(2) if $a \in R$ and $X \in V$, then $n(aX) = |a|n(X)$;
(3) $n(X + Y) \leq n(X) + n(Y)$ for all $X,\ Y \in V$.

We shall also use the notation

$$n(X) = |X| = \|X\|.$$

In particular, if B is a positive definite symmetric bilinear form on the vector space V over R, then

$$\|X\| = B(X, X)^{1/2}$$

is a norm on V and we have the inequality

$$|B(X, Y)| \leq \|X\|\|Y\|.$$

Using a norm on V we can define a **metric** d on V by

$$d(X, Y) = \|X - Y\|$$

for $X,\ Y \in V$. Thus d satisfies

(1) $d(X, Y) \geq 0$ and $d(X, Y) = 0$ if and only if $X = Y$;
(2) $d(X, Y) = d(Y, X)$;
(3) $d(X, Y) \leq d(X, Z) + d(Z, Y)$.

In particular, with $\|X\| = B(X,\ X)^{1/2} = (\sum x_i{}^2)^{1/2}$ we obtain $d(X,\ Y) = [\sum(x_i - y_i)^2]^{1/2}$.

We now consider some of the topological properties of $V = R^n$ which arise from a metric d obtained from a given fixed norm. Thus we define the **open ball of radius** r **with center** p by

$$B(p, r) = \{X \in V : d(p, X) < r\}$$

and say subset S of V is **open** in V if for every $p \in S$ there exists $r > 0$ so that the open ball $B(p, r)$ is contained in S. Using this definition we obtain the basic results on the metric topology of R^n with which we assume the reader is familiar.

Notice that it really does not matter which norm we start with when considering the topological properties of $V = R^n$ relative to a metric induced by a norm. Thus if n_1 and n_2 are norms on V, we can show that there exist constants a and b in R so that for all $X \in V$

$$an_1(X) \leq n_2(X) \leq bn_1(X);$$

that is, n_1 and n_2 are equivalent norms. Thus, if d_k is the metric determined by the norm n_k, then using the above inequality it is easy to see that the open sets of V relative to d_1 are the same as the open sets relative to d_2 [Dieudonné, 1960; Lang, 1968]. Most often we shall determine a norm n on V by choosing a basis $X_1, \ldots, X_n$ of V and for $X = \sum x_i X_i$ set $n(X) = (\sum x_i^2)^{1/2}$. In particular we obtain the usual inner product and norm by taking the basis $e_1, \ldots, e_n$ of $V = R^n$ and call the vector space R^n with the topology induced by this norm **Euclidean n-space.**

With the topology in $V = R^n$ induced by a norm n as above we now note that V is complete. Thus a sequence of vectors $\{x_k\}$ in V is called a **Cauchy sequence** if given any $\varepsilon > 0$ there exists N so that for all p, $q \geq N$ we have

$$n(x_p - x_q) < \varepsilon.$$

We have the result that every Cauchy sequence in V has a limit; that is, V is complete [Dieudonné, 1960; Lang, 1968]. Let $\{x_k\}$ be a sequence in V and let $X_1, \ldots, X_n$ be a basis of V. Then we can write

$$x_k = x_{k1} X_1 + \cdots + x_{kn} X_n$$

and note $\{x_k\}$ converges if and only if each sequence $\{x_{ki}\}$, $i = 1, \ldots, n$, converges. Thus, by a skillful choice of a basis, it might be obvious that the sequences $\{x_{ki}\}$ converge so that $\{x_k\}$ can be shown to converge easily.

Let $\{x_k\}$ be a sequence in $V = R^n$. Then the series of vectors $\sum x_k$ in V converges if the sequence $\{s_p\}$ given by the partial sums

$$s_p = \sum_{k=1}^{p} x_k$$

converges. Now associated with any series $\sum x_k$ in V is the series of real numbers

$$\sum n(x_k)$$

formed by taking the series of norms of each term. We have the following expected results [Dieudonné, 1960; Lang, 1968].

(1) If the series $\sum n(x_k)$ converges in R, then the series $\sum x_k$ converges in the Euclidean space R^n; in this case we say $\sum x_k$ **converges absolutely.**

(2) If $\sum x_k$ converges absolutely to the limit a, then the series obtained by any rearrangement of the terms also converges absolutely to a.

Let $V = R^n$ and $W = R^m$ be Euclidean spaces as previously discussed and let U be a nonempty subset of V. Then for a choice of basis $Y_1, \ldots, Y_m$ of W a function

$$f : U \to W : p \to f(p)$$

has the coordinate representation

$$f(p) = \sum_{k=1}^{m} f_k(p) Y_k .$$

The **coordinate functions** for f

$$f_k : U \to R : p \to f_k(p)$$

can be used to describe the boundedness or continuity of the function $f : U \to W$ as follows.

A function $f : U \to W$ is **bounded** on U if there exists a real number $C > 0$ so that $n(f(u)) \leq C$ for all $u \in U$, where n is the Euclidean norm in W. If we let

$$\mathbf{B}(U, W)$$

denote the set of bounded functions on U into W, then a straightforward computation shows $\mathbf{B}(U, W)$ is a vector space of functions over R relative to the usual operations

$$(af + bg)(p) = af(p) + bg(p).$$

Furthermore it is easy to see that $f \in \mathbf{B}(U, W)$ if and only if each coordinate function $f_k \in \mathbf{B}(U, R)$.

Similarly with the usual definition of a **continuous function** $f : U \to W$ we find that the set of all such functions is a vector space of functions over R which we denote by

$$C(U, W).$$

In particular, for a function $f : U \to W$, we have that $f \in C(U, W)$ if and only if each coordinate function $f_k \in C(U, R)$. We shall frequently use the notation

$$C(U) = C(U, R)$$

and note that $C(U)$ is closed under the pointwise product

$$(fg)(p) = f(p)g(p).$$

Thus $C(U)$ becomes an associative algebra relative to this product.

The vector space of bounded functions $\mathbf{B}(U, W)$ has a **sup norm** given by

$$\|f\| = \sup\{\|f(x)\| : x \in U\},$$

and the uniform convergence of a sequence $\{f_n\}$ of functions

$$f_n : U \to W$$

is related to the norm as follows. Recall $\{f_n\}$ converges uniformly on U to a function $f : U \to W$ if for every $\varepsilon > 0$ there is an integer N such that $n \geq N$ implies for all $x \in U$

$$\|f_n(x) - f(x)\| < \varepsilon.$$

Consequently in terms of norms, $\{f_n\}$ converges uniformly to f on U if for every $\varepsilon > 0$ there is an integer N such that for all $n \geq N, f_n - f \in \mathbf{B}(U, W)$ and

$$\|f_n - f\| < \varepsilon.$$

Using this together with the fact that if U is compact, then $C(U) \subset \mathbf{B}(U, R)$ we obtain the following summary of results [Dieudonné, 1960; Lang, 1968]:

Theorem Let V, W be Euclidean spaces and let U be a nonempty subset of V.

(a) The space $\mathbf{B}(U, W)$ with sup norm is complete.

(b) If $\{f_n\}$ is a sequence of functions in $C(U, W)$ which converges uniformly on U to a function $f : U \to W$, then $f \in C(U, W)$.

(c) The space $BC(U, W)$ of bounded continuous functions from U into W is complete and closed in $\mathbf{B}(U, W)$ (in the sup norm). In particular, if U is compact, this applies to $C(U) = BC(U, R)$.

(d) (Weierstrass test) Let $\{f_n\}$ be a sequence in $\mathbf{B}(U, W)$ so that there exist real numbers M_n with $\|f_n\| \leq M_n$ and $\sum M_n$ convergent. Then the series $\sum f_n$ converges absolutely and uniformly. Furthermore if each f_n is continuous on U, then $\sum f_n$ is continuous on U.

Remarks (1) If we use the field $K = C$, the complex numbers, then results analogous to those discussed in this section also hold. However, we must use a **Hermitian form** instead of the inner product to define the metric. Thus if $V = C^n$, a Hermitian form is a mapping $H : V \times V \to C$ satisfying for $X, Y, Z \in V$ and $a \in C$,

(i) $H(X, Y + Z) = H(X, Y) + H(X, Z)$;

(ii) $H(aX, Y) = aH(X, Y)$ and $H(X, aY) = \bar{a}H(X, Y)$ where the overbar denotes the complex conjugate;

(iii) $H(X, Y) = \overline{H(Y, X)}$;

(iv) $H(X, X) > 0$ if $X \neq 0$ and $H(X, X) = 0$ if and only if $X = 0$.

Note that from (iii) $H(X, X) = \overline{H(X, X)}$ so that $H(X, X) \in R$. Thus (iv) allows us to define a norm by $\|X\| = H(X, X)^{1/2}$ and consequently a metric. In particular, for $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ in $V = C^n$, $H(X, Y) = \sum x_k \bar{y}_k$ defines a Hermitian form and $\|X\| = (\sum x_k \bar{x}_k)^{1/2}$.

(2) We use the following convention to obtain a matrix for

$$T \in \operatorname{Hom}(V, W).$$

Let $X_1, \ldots, X_n$ be a basis of V and let $Y_1, \ldots, Y_m$ be a basis of W and let $T(X_i) = \sum_{j=1}^{m} a_{ji} Y_j$. Then the matrix for T relative to these bases is (a_{ji}); that is, the ith column is the set of coefficients obtained from the above expression for $T(X_i)$. In particular, the various canonical forms will have upper triangular matrices.

Exercises (1) Let $X_1, \ldots, X_n$ be a basis of V over C so that any $T \in \operatorname{End}(V) = L(V, V)$ has matrix (t_{ij}) relative to this basis. Show

$$\|T\| = \sup\{|t_{ij}|\}$$

defines a norm on End(V). What happens when one changes basis?

(2) For $T \in \operatorname{End}(V)$ show that

$$\|T\| = \sup\{\|TX\|/\|X\| : X \neq 0\}$$

defines a norm on End(V) such that

$$\|T\| \geq \max\{|r_1|, \ldots, |r_n|\}$$

where the r_i are the characteristic roots of T.

Example (1) We now consider some series in End(V). First we note that for $T \in \operatorname{End}(V)$ the sup norm

$$\|T\| = \sup\{\|T(X)\|/\|X\| : X \neq 0\} = \sup\{\|T(Y)\| : \|Y\| = 1\}$$

satisfies, for S, $T \in \operatorname{End}(V)$,

$$\|ST\| \leq \|S\| \|T\|.$$

In particular, $\|T^n\| \leq \|T\|^n$. Next in discussing the convergence of a series $\sum a_n T^n$ in End(V) we make a choice of basis in V for which the matrix T has a desirable form. Then we can define a norm on V induced by this basis which can be used to compute the sup norm on End(V) in terms of the matrix for T. Thus if T is represented by matrices A and B relative to different bases so that $B = PAP^{-1}$, then using the norms defined by these basis we see $\sum a_n A^n$ exists if and only if $\sum a_n B^n$ exists. This is because

$$\sum a_n B^n = \sum a_n (PAP^{-1})^n = P(\sum a_n A^n)P^{-1}.$$

Now note that we can regard the real vector space V as contained in a complex vector space W over C so that the norm in V is induced by a norm in W given by a Hermitian form. Thus we can use the complex canonical forms for the matrix of T to investigate convergence of $\sum a_n T^n$. We use this now to sketch a proof of the following results [Jacobson, 1953, Vol. II].

Proposition Let $\sum a_j z^j$ be a power series over the complex numbers C with radius of convergence ρ and let $T \in \operatorname{End}(W)$ be such that its characteristic roots r_k satisfy $|r_k| < \rho$. Then the power series $\sum a_j T^j$ exists in End(W). In particular this result holds for real power series.

Proof Choose a basis of W so that the matrix A of T has canonical form

$$A = \begin{bmatrix} A_1 & & & & \\ & A_2 & & & 0 \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & \cdot \\ & & & & & A_p \end{bmatrix}$$

where each block has the form

$$\begin{bmatrix} r & 1 & & & 0 \\ & \cdot & & & \\ & r & & \cdot & \\ & & \cdot & & 1 \\ & & & \cdot & \\ 0 & & & & r \end{bmatrix} \qquad (*)$$

where r is a characteristic root of T. Consequently, since the matrix of the powers T^k have the same block form, it suffices to consider the series for a matrix A of the form $(*)$. Now if

$$s_k(z) = \sum_{j=0}^{k} a_j z^j, \qquad s_k'(z) = \sum_{j=0}^{k} j a_j z^{j-1}, \qquad \text{etc.},$$

then the kth partial sum $s_k(A)$ for the series is of the form

$$\begin{bmatrix} s_k(r) & s_k'(r) & \dfrac{s_k''(r)}{2!} & \\ & s_k(r) & s_k'(r) & \\ 0 & & s_k(r) & \\ & & & \ddots \end{bmatrix}.$$

Thus if $|r| < \rho$, then the sequences $\{s_k(r)\}$, $\{s_k'(r)\}$, etc., converge to $s(r)$, $s'(r)$, etc., so that the sequence $\{s_k(A)\}$ converges; that is, $\sum a_k T^k$ exists.

We apply this to the complex power series

$$e^z = 1 + z + \frac{1}{2!} z^2 + \cdots$$

which converges for all $z \in C$. Thus for any $T \in \mathrm{End}(V)$ we note that the characteristic roots of T are bounded so that

$$\exp(T) = e^T = I + T + \frac{1}{2!} T^2 + \cdots$$

exists. Thus we have the mapping

$$\exp : \mathrm{End}(V) \to \mathrm{End}(V) : T \to e^T$$

called the **exponential mapping** (for endomorphisms).

If T has a matrix A of the form $(*)$ above, then $\exp A$ is a matrix of the form

$$\begin{bmatrix} \exp r & \exp r & \dfrac{\exp r}{2!} & \\ & \exp r & \exp r & \\ 0 & & \exp r & \\ & & & \ddots \end{bmatrix}.$$

Thus in general $\exp T$ is computed from blocks of such matrices.

Exercises Let $V = R^n$ be Euclidean n-space and let $T \in \mathrm{End}(V)$.

(3) Using the above canonical form for $\exp T$, show $\det(\exp T) = e^{tr\,T}$, where $tr\,T = \mathrm{trace}\,T$ is the sum of the characteristic roots of T. Thus the exponential map has domain $\mathrm{End}(V)$ and range contained in $GL(V)$ the group of nonsingular endomorphisms of V. We also use the notation $GL(n, R)$ for $GL(V)$.

(4) Show that the series for $\exp T$ is absolutely and uniformly convergent on any closed ball in $\mathrm{End}(V)$. (If possible, try not to use the Weierstrass test but estimate directly). Thus show $\exp : \mathrm{End}(V) \to GL(V)$ is continuous.

(5) Show $GL(V)$ is an open subset of the Euclidean space $\mathrm{End}(V)$.

2. The Derivative

In this section we formulate the basic facts on differentiation in terms of linear transformations. We do not give many proofs so the student should regard many of the statements as exercises or refer to Dieudonné [1960], Lang [1968], or Spivak [1965].

Definition 1.1 Let V and W denote the Euclidean spaces R^n and R^m, respectively, and let U be an open subset of V. Then the function

$$f : U \to W$$

is **differentiable** at $p \in U$ if there exists a linear transformation $T \in L(V, W)$ so that for all $X \in V$ the limit

$$\lim_{X \to 0} \frac{\|f(p+X) - f(p) - T(X)\|}{\|X\|} = 0,$$

where $\| \ \|$ denotes the usual Euclidean norm in V or W.

Thus for X sufficiently small in V we have the approximation for a differentiable function

$$f(p+X) = f(p) + T(X) + \|X\|\varepsilon(X), \tag{$*$}$$

where $\varepsilon(X)$ is a function such that

$$\lim_{X \to 0} \varepsilon(X) = 0.$$

If such a linear transformation T exists, then T is called **the derivative** of f at $p \in U$ and is denoted by $Df(p)$, $f'(p)$, or $df(p)$. Thus

$$Df(p)X = T(X)$$

for all $X \in V$. This definition uses the uniqueness part of the following result.

Proposition 1.2 Let $f : U \to W$ be differentiable at $p \in U$ and let $T_1, T_2 \in L(V, W)$ satisfy

$$\lim_{X \to 0} \frac{\|f(p+X) - f(p) - T_i(X)\|}{\|X\|} = 0$$

for $i = 1, 2$ and all $X \in V$. Then $T_1 = T_2 \equiv Df(p)$. Furthermore for all $X \in V$

$$Df(p)(X) = \lim_{t \to 0} \frac{1}{t} [f(p+tX) - f(p)].$$

PROOF For $0 \neq X \in V$ let $d(X) = f(p+X) - f(p)$. Then we have

$$\begin{aligned}
\frac{\|T_1(X) - T_2(X)\|}{\|X\|} &= \lim_{t \to 0} \frac{\|T_1(tX) - T_2(tX)\|}{\|tX\|} \\
&= \lim_{t \to 0} \frac{\|T_1(tX) - d(tX) + d(tX) - T_2(tX)\|}{\|tX\|} \\
&\leq \lim_{t \to 0} \|\varepsilon_1(tX) + \varepsilon_2(tX)\| = 0,
\end{aligned}$$

where we use the approximation (∗). This implies $T_1(X) = T_2(X)$ for all $X \in V$. Using (∗) we also have

$$Df(p)(X) = \lim_{t\to 0} \frac{1}{t} Df(p)(tX) + \lim_{t\to 0} \frac{1}{t} \|tX\|\varepsilon(tX)$$

$$= \lim_{t\to 0} \frac{1}{t} [f(p + tX) - f(p)].$$

Examples (1) Let $f : V \to W$ be a linear transformation. Then $Df(p) = f$ for all $p \in V$. For let $X \in V$, then $f(p + X) - f(p) = f(X)$. Thus for $T = f$ we see $f(p + X) - f(p) - T(X) = 0$ so that the limit in the definition is 0 and by uniqueness $Df(p) = f$.

(2) Let $V = W = R$ and let $f : U \to W$ be differentiable at p in the usual calculus sense with

$$f'(p) = \lim_{t\to 0} \frac{1}{t} [f(p + t) - f(p)].$$

Then the linear transformation $T : R \to R : X \to f'(t)X$ given by multiplication satisfies Definition 1.1. Thus $Df(p)$ has the 1×1 matrix $(f'(p))$ relative to the basis of R consisting of the number 1.

Proposition 1.3 Let V and W be Euclidean spaces with U an open subset of V.

(a) If $f : U \to W$ is differentiable at $p \in U$, then f is continuous at p.

(b) If $f, g : U \to W$ are functions differentiable at $p \in U$ and if $a, b \in R$, then $af + bg$ is differentiable at p and

$$D(af + bg)(p) = aDf(p) + bDg(p).$$

(c) Let V_1, V_2 be Euclidean spaces and let $B : V_1 \times V_2 \to W$ be a bilinear map. Suppose $f : U \to V_1$ and $g : U \to V_2$ are differentiable at $p \in U$. Then the "product" $B(f, g) : U \to W: u \to B(f(u), g(u))$ is differentiable at p and for $X \in V$,

$$[D(B(f, g))(p)](X) = B(f'(p)X, g(p)) + B(f(p), g'(p)X).$$

(d) (Chain rule) Suppose U_1 is open in V_1, U_2 is open in V_2, and $f : U_1 \to U_2$ and $g : U_2 \to W$. Let p be an element of U_1 such that f is differentiable at p and g is differentiable at $f(p)$. Then the composition $g \circ f : U_1 \to W$ is differentiable at p and

$$D(g \circ f)(p) = Dg(f(p)) \circ Df(p).$$

Thus the derivative of the composition is just the composition of the derivatives regarded as linear transformations.

Proof of (d) Let $0 \neq X \in V_1$ be near 0 in V_1, let $S = Df(p)$, and let $T = (Dg)(f(p))$. Then

$$f(p + X) = f(p) + Y$$

where $Y = Df(p)X + \|X\|\varepsilon(X)$ and

$$g(f(p + X)) = g(f(p) + Y) = g(f(p)) + [Dg(f(p))]Y + \|Y\|\varepsilon_1(Y).$$

Let $\Delta(X) = (g \circ f)(p + X) - (g \circ f)(p) - T \circ S(X)$. Then

$$\begin{aligned}\|\Delta(X)\|/\|X\| &= \|g(f(p + X)) - g(f(p)) - TS(X)\|/\|X\| \\ &= \|[Dg(f(p))]Y + \|Y\|\varepsilon_1(Y) - TS(X)\|/\|X\| \\ &= \|[Dg(f(p))](\|X\|\varepsilon(X)) + \|Y\|\varepsilon_1(Y)\|/\|X\|.\end{aligned}$$

Thus for $M = \|Dg(f(p))\|$ and $N = \|Df(p)\|$, the norm of a linear transformation as in Section 1.1, we have

$$\begin{aligned}\|Y\| &= \|Df(p)X + \|X\|\varepsilon(X)\| \\ &\leq \|Df(p)X\| + \|X\|\|\varepsilon(X)\| \leq (N + \|\varepsilon(X)\|)\|X\|.\end{aligned}$$

Consequently

$$\|\Delta(X)\|/\|X\| \leq M\|\varepsilon(X)\| + (N + \|\varepsilon(X)\|)\|\varepsilon_1(Y)\|$$

so that

$$\lim_{X \to 0} \|\Delta(X)\|/\|X\| = 0$$

which yields $g \circ f$ differentiable at p and $D(g \circ f)(p) = T \circ S$.

Examples (3) Let $B : V_1 \times V_1 \to R$ be a bilinear form and let

$$U = (-1, 1) \subset V = R.$$

For fixed $Y \in V_1$, let

$$f : U \to V_1 : t \to A(t)Y.$$

where $A(t) \in \operatorname{End}(V_1)$ is such that $A : t \to A(t)$ is differentiable on $(-1, 1)$ and $A(0) = I$. Thus $A(t)$ has a matrix $(a_{ij}(t))$ such that $a_{ij}(0) = \delta_{ij}$ and the a_{ij} are differentiable functions on $(-1, 1)$ to R. This gives a map

$$g : U \to R : t \to B(f(t), f(t)) = B(A(t)Y, A(t)Y),$$

and for $p = 0 \in U$ and $X = 1 \in V$ in the formula for the product rule we have

$$\begin{aligned}[Dg(0)](1) &= [D(B(f, f))(0)](1) \\ &= B(f'(0)1, f(0)) + B(f(0), f'(0)1) \\ &= B(A'(0)Y, Y) + B(Y, A'(0)Y),\end{aligned}$$

where, with the notation $da_{ij}/dt(0) = a'_{ij}(0)$, $A'(0)$ has the matrix

$$(da_{ij}/dt(0)).$$

In particular let $A(t)$ be a B-isometry; that is, for all $Y \in V_1$

$$B(A(t)Y, A(t)Y) = B(Y, Y).$$

Then since the derivative of a constant function is 0, we obtain

$$B(A'(0)Y, Y) + B(Y, A'(0)Y) = 0;$$

that is, $A'(0)$ is skew-symmetric relative to B.

(4) Let $f : R \to R$ and $g : R \to R$ be differentiable functions such that $g \circ f = idy$, the identity function on R. Then from the chain rule, $idy = Dg(f(p)) \circ Df(p)$ and applying this to the vector $1 \in R$ we obtain

$$1 = Dg(f(p)) \cdot Df(p)(1) = g'(f(p)) \cdot f'(p)$$

which gives the usual formula for the derivative of the inverse function.

In order to compute a matrix for the derivative $Df(p)$ of a differentiable function $f : U \to W$ we consider a coordinate representation for f. First let

$$W = W_1 \times W_2 \times \cdots \times W_k$$

be a product of Euclidean spaces. Then $f : U \to W$ yields the "coordinate" functions $f_i : U \to W_i$ by

$$f(p) = (f_1(p), \ldots, f_k(p))$$

and we leave the following as an exercise.

Proposition 1.4 Let U be an open subset of V and let

$$f : U \to W_1 \times \cdots \times W_k$$

be given by $f = (f_1, \ldots, f_k)$ as above. Then f is differentiable at $p \in U$ if and only if each coordinate map f_i is differentiable at p; in this case

$$Df(p) = (Df_1(p), \ldots, Df_k(p)).$$

In particular for $Y_1, \ldots, Y_m$ a basis of W and $f = \sum f_j Y_j$ we have for $X \in V$

$$Df(p)(X) = \sum_j [Df_j(p)(X)]Y_j.$$

Next we use a factorization

$$V = V_1 \times V_2 \times \cdots \times V_r$$

as a product of Euclidean spaces and consider partial derivatives. Thus let $x = (x_1, \ldots, x_r) \in V$ be written in terms of its "coordinates" and for $p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_r$ fixed we consider the map of an open set $U_i \subset V_i$

$$f_i : U_i \to W : x_i \to f(p_1, \ldots, x_i, \ldots, p_r).$$

If f_i is differentiable at $p_i \in U_i$, we call its derivative the ith **partial derivative of** f at

$$p = (p_1, \ldots, p_i, \ldots, p_r) \in U_1 \times \cdots \times U_i \times \cdots \times U_r$$

and denote it by $D_i f(p)$. Thus if this partial derivative exists, we have for $X_i \in V_i$ that

$$\begin{aligned} D_i f(p)(X_i) &= \lim_{t \to 0} \frac{1}{t} [f(p_1, \ldots, p_i + tX_i, \ldots, p_r) - f(p_1, \ldots, p_i, \ldots, p_r)] \\ &= Df(p)(0, \ldots, X_i, 0, \ldots, 0). \end{aligned}$$

In particular for $W = R$ and $V = R^n = R \times \cdots \times R$ n-times, we use the familiar notation

$$\partial f / \partial x_i(p) = D_i f(p)(e_i) = Df(p)(e_i),$$

where $e_i = (0, \ldots, 1, 0, \ldots, 0)$ with 1 in the ith position. However, "partial derivatives" can be defined relative to any basis of R^n.

Proposition 1.5 Let $V = V_1 \times \cdots \times V_r$ be a product of Euclidean spaces and let U_i be an open subset of V_i. Let W be a Euclidean space and let

$$f : U_1 \times \cdots \times U_r \to W$$

be a function differentiable at $p = (p_1, \ldots, p_r) \in U_1 \times \cdots \times U_r$. Then each partial derivative $D_i f(p)$ exists and for $X = (X_1, \ldots, X_r) \in V_1 \times \cdots \times V_r$ we have

$$Df(p)X = \sum_{i=1}^{m} D_i f(p) X_i.$$

PROOF We leave the proof to the reader as an exercise but the method should be familiar: Either adding and subtracting the same suitable terms from $f(p + X) - f(p)$, or showing that $Df(p)$ exists implies $D_i f(p)$ exists, then using the linearity of $Df(p)$ on $X = \sum(0, \ldots, X_i, 0, \ldots, 0)$ and the corresponding formula for $D_i f(p) X_i$.

In particular note that if $W = R$ and $V = R \times \cdots \times R$ with

$$X = (x_1, \ldots, x_n) = \sum x_i e_i \in V,$$

then we have

$$Df(p)X = \sum D_i f(p)(x_i e_i) = \sum x_i \, \partial f / \partial x_i(p).$$

We now put together the above results to obtain a matrix for $Df(p)$.

Theorem 1.6 Let V and W be Euclidean spaces with $X_1, \ldots, X_n$ a basis of V and $Y_1, \ldots, Y_m$ a basis of W. Let U be an open subset of V and let

$$f : U \to W$$

be a function which is differentiable at $p \in U$ and is given by the coordinate functions

$$f(x) = \sum_{j=1}^{m} f_j(x) Y_j$$

where $x = \sum_{i=1}^{n} x_i X_i \in U$. Then $Df(p)$ has the matrix

$$(D_i f_j(p)(X_i))$$

relative to the basis $X_1, \ldots, X_n$ of V and $Y_1, \ldots, Y_m$ of W; this matrix is called a **Jacobian matrix.** In case we choose the basis $e_1, \ldots, e_n$ of V and $e_1, \ldots, e_m$ of W, then $Df(p)$ has the matrix

$$(\partial f_j / \partial x_i(p)).$$

Proof For a typical basis element X_i we shall show

$$Df(p)X_i = \sum_j [Df_j(p)(X_i)] Y_j$$

which gives the desired matrix. Recall the convention that a linear transformation $T : V \to W$ has matrix (a_{ji}) relative to the above bases provided $T(X_i) = \sum_j a_{ji} Y_j$.

From Proposition 1.4 we have each coordinate function f_j is differentiable and

$$Df(p)(X_i) = \sum_j [Df_j(p)(X_i)] Y_j = \sum_j [D_i f_j(p)(X_i)] Y_j$$

using Proposition 1.5 concerning the partial derivative $D_i f_j(p)$ evaluated at X_i.

Remarks (1) A more basic proof using the approximation (*) after Definition 1.1 is as follows.

$$\begin{aligned} Df(p)(X_i) - \sum_j Df_j(p)(X_i) Y_j &= \lim_{t \to 0} \frac{1}{t} [f(p + tX_i) - f(p) - \sum_j Df_j(p)(tX_i) Y_j] \\ &= \lim_{t \to 0} \frac{1}{t} \sum_j [f_j(p + tX_i) - f_j(p) - Df_j(p)(tX_i)] Y_j \\ &= \lim_{t \to 0} \frac{1}{t} \sum_j \|tX_i\| \varepsilon_j(tX_i) Y_j = 0. \end{aligned}$$

(2) The results of Propositions 1.4 and 1.5 can be used to find a Jacobian matrix for $Df(p)$ in "block" form. Thus let $V = V_1 \times V_2$ and $W = W_1 \times W_2$ be factorizations into Euclidean subspaces and let U be open in V. Let the function $f : U \to W$ be given by the coordinate functions

$$f_1 : U \to W_1 \qquad \text{and} \qquad f_2 : U \to W_2.$$

Then if f is differentiable at $p \in U$, the linear transformation $Df(p)$ has the matrix

$$\begin{bmatrix} D_1 f_1(p) & D_2 f_1(p) \\ D_1 f_2(p) & D_2 f_2(p) \end{bmatrix},$$

where the $D_r f_s(p)$ are represented by appropriate matrices. Note that some texts have the transpose of the above matrix for their derivative. This depends only on summation conventions and we shall write this in detail when necessary.

Definition 1.7 Let U be open in V and let $f : U \to W$ be differentiable at every point $p \in U$. Then f is **continuously differentiable on** U or of **class** C^1 on U if the map

$$U \to \operatorname{Hom}(V, W) : p \to Df(p)$$

is continuous. We denote the set of these continuously differentiable functions by

$$C^1(U, W)$$

and for $W = R$ we use the notation

$$C^1(U) = C^1(U, R).$$

A straightforward computation shows $C^1(U, W)$ is a vector space and a differentiable function $f : U \to W$ is in $C^1(U, W)$ if and only if we have a matrix representation $(a_{ij}(p))$ of $Df(p)$ with all $a_{ij} : U \to R : p \to a_{ij}(p)$ continuous functions.

Generalizing the results concerning term-by-term differentiation of series we have the following result (also see Dieudonné [1960]).

Proposition 1.8 Let V and W be Euclidean spaces with U open in V and let $\{f_n\}$ be a sequence of functions in $C^1(U, W)$ and let $\{Df_n\}$ be the corresponding sequence of derivatives. Assume there exists a function $f : U \to W$ so that for all $x \in U$

$$f(x) = \lim_{n \to \infty} f_n(x)$$

with pointwise convergence and assume that there is a function $g : U \to \operatorname{Hom}(V, W)$ so that

$$Df_n \to g$$

with uniform convergence on U. Then f is differentiable on U and $Df(p) = g(p)$ for all $p \in U$.

Exercises (1) Show the map

$$f : GL(V) \to GL(V) : T \to T^{-1}$$

is differentiable at P and

$$Df(P)T = -P^{-1}TP^{-1}.$$

Thus show $f \in C^1(GL(V), \operatorname{End}(V))$.

(2) (i) For $T \in \operatorname{End}(V)$ show the function

$$\phi : R \to \operatorname{End}(V) : t \to \exp(tT)$$

is differentiable at any point $p \in R$.

(ii) Show the exponential function

$$\exp : \operatorname{End}(V) \to GL(V) : T \to e^T$$

is of class C^1 on a suitable neighborhood U of 0 in $\operatorname{End}(V)$.

(3) Prove the following version of the **mean value theorem.** Let U be open in V and let $f : U \to W$ be of class C^1 on U. Let $p \in U$ and $X \in V$ be such that the line segment $p + tX$ for $0 \le t \le 1$ is contained in U. Then

$$\|f(p + X) - f(p)\| \le \sup\|Df(q)\|\,\|X\|,$$

where the sup is taken over all q on the line segment.

(4) Let $B : R^2 \times R^2 \to R$ be a bilinear form given by

$$B((x_1, x_2), (y_1, y_2)) = x_1y_2 - x_2y_1.$$

Relative to the usual basis of R^2 let the following represent endomorphisms of R^2

$$A(t) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, \qquad B(t) = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},$$

$$C(t) = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}, \qquad D(t) = A(at)B(bt)C(ct)$$

for a, b, $c \in R$, and t in a suitable interval in R.

(i) Show $A(t)$, $B(t)$, $C(t)$, $D(t)$ are B-isometries [see example (3), Section 1.2].

(ii) Compute $D'(0)$.

3. Higher Derivatives

Let U be open in $V = R^n$ and let $f: U \to W$ be differentiable on U. Then

$$Df: U \to L(V, W): p \to Df(p)$$

is a function defined on U with values in the vector space $L(V, W) = \mathrm{Hom}(V, W)$. Thus if the appropriate limits exist, we can define the **second derivative** at $p \in U$ by

$$D^2f(p) = D[Df](p) \in L(V, L(V, W)).$$

Next let $L^2(V, W)$ denote the space of bilinear maps from $V \times V$ into W. Then we can identify the elements of $L(V, L(V, W))$ with those of $L^2(V, W)$ as follows. Let X, $Y \in V$. Then for $B \in L(V, L(V, W))$ we see that $B(X) \in L(V, W)$. Thus $[B(X)](Y) \equiv B'(X, Y) \in W$ is linear in X and Y and therefore $B' \in L^2(V, W)$. From this the mapping $B \to B'$ is an isomorphism of the above spaces but we shall consistently use the above identification. In particular for a function $f: U \to W$ differentiable on U such that $B = D^2f(p)$ exists for $p \in U$, let

$$F(p) = Df(p) \in L(V, W).$$

Then for X, $Y \in V$ we have

$$D^2f(p)(X, Y) = [DF(p)(X)]Y.$$

Example (1) Let $f: U \to R$ be differentiable on $U \subset V = R^n$ and suppose $D^2f(p)$ exists. Then $D^2f(p): V \times V \to R$ is a bilinear form and we shall now compute its matrix. Let $e_1, \ldots, e_n$ be the usual basis of $V = R^n$ and let $g_i(u) = Df(u)(e_i)$; that is,

$$g_i(u) = D_i f(u)(e_i) = (\partial f/\partial x_i)(u).$$

Then $F(u) = Df(u)$ is given by

$$F(u)Z = Df(u)Z = \textstyle\sum z_i\, Df(u)(e_i) = \sum z_i g_i(u), \tag{$*$}$$

where $Z = \sum z_i e_i \in V$. Next we note for $X = \sum x_j e_j \in V$ that

$$Dg_i(p)(X) = \sum_j x_j Dg_i(p)(e_j) = \sum_j x_j D_j D_i f(p)(e_i, e_j). \tag{$**$}$$

Thus for $Y = \sum y_k e_k$,

$$\begin{aligned} D^2 f(p)(X, Y) &= [DF(p)(X)]Y \\ &= \lim_{t \to 0} \frac{1}{t} [F(p + tX)Y - F(p)Y] \\ &= \lim_{t \to 0} \frac{1}{t} \sum_k y_k [g_k(p + tX) - g_k(p)] \\ &= \sum_k y_k Dg_k(p)(X) \\ &= \sum_k \sum_j x_j y_k D_j D_k f(p)(e_k, e_j), \end{aligned}$$

using $(*)$ for the third equality and $(**)$ for the fifth equality. Thus with the notation

$$\partial^2 f / \partial x_i \, \partial x_j(p) = D_i D_j f(p)(e_j, e_i)$$

we see that the matrix for the bilinear form $D^2 f(p)$ is the **Hessian matrix**

$$\big(\partial^2 f / \partial x_i \, \partial x_j(p)\big).$$

For the general case of a differentiable function $f : U \to W$ we can find a formula for $D^2 f(p)$ (assuming it exists) by putting together the Hessian matrices for the coordinate functions. Thus let $\bar{e}_1, \ldots, \bar{e}_m$ be the usual basis for $W = R^m$. Then writing $f(u) = \sum f_i(u)\bar{e}_i$, where $f_i : U \to R$, we have for $X, Y \in V$

$$D^2 f(p)(X, Y) = \sum_{i=1}^{m} D^2 f_i(p)(X, Y)\bar{e}_i$$

noting D^2 is linear and $D^2 f_i(p)$ exist.

Exercise (1) Let $f : R \to R$ be given by $f(x) = \cos 2x$. Relate the "usual" second derivative of f at p to the second derivative discussed above.

Definition 1.9 A function $f : U \to W$ of class C^1 on U is of **class C^2 on U** if for all $p \in U$, $D^2 f(p)$ exists and the mapping

$$U \to L^2(V, W) : p \to D^2 f(p)$$

is continuous on U.

It is easy to see that the set of all functions $f : U \to W$ of class C^2 on U is a vector space and when $W = R$ we frequently denote this vector space by $C^2(U)$. Thus $f \in C^2(U)$ if and only if all the partial derivatives of order less than or equal to 2 exist and are continuous on U. In this case it is known that we can interchange the order of differentiation and more generally we have the following result. (For the proof see Dieudonné [1960] and Lang [1968].)

Proposition 1.10 Let U be open in V and let $f : U \to W$ be of class C^2 on U. Then for all $p \in U$ and $X, Y \in V$,

$$D^2f(p)(X, Y) = D^2f(p)(Y, X).$$

Thus $D^2f(p)$ is a symmetric bilinear form on $V \times V$ into W.

Thus we see that if $f \in C^2(U)$, then the Hessian matrix

$$\left(\partial^2 f/\partial x_i \, \partial x_j(p)\right)$$

is actually a symmetric matrix.

We now define higher-order derivatives by induction. Thus for $f : U \to W$ and $p \in U$ we set

$$D^r f(p) = D[D^{r-1}f](p)$$

if the appropriate limit exists. The rth derivative

$$D^r f(p) \in L\big(V, L(V, \ldots, L(V, W) \ldots)\big)$$

but, as in the case $r = 2$, we can identify this with the set of multilinear functionals from $V^r = V \times \cdots \times V$ to W which we denote by $L^r(V, W)$.

Definition 1.11 (a) Let U be open in V and let $f : U \to W$ be continuous on U. Then f is of **class C^p on** U or a **C^p function** if for all $u \in U$, $D^r f(u)$ exists and if

$$D^r f : U \to L^r(V, W) : u \to D^r f(u)$$

is continuous for $r = 1, \ldots, p$. We use the notation $C^p(U)$ for the functions of class C^p when $W = R$.

(b) A function $f : U \to W$ is of **class** C^∞ on U or **infinitely differentiable** if f is of class C^r for all $r = 1, 2, \ldots$. When $W = R$ we use the notation $C^\infty(U)$ for the class of infinitely differentiable functions.

The following summarizes the results on C^r-functions and for proofs see Dieudonné, [1960] and Lang [1968].

Theorem 1.12 Let U be open in $V = R^n$.

(a) If $f : U \to W$ is of class C^r on U, then for each $p \in U$ we have $D^r f(p)$ is a symmetric r-multilinear form; that is,

$$D^r f(p)(X_1, \ldots, X_r) = D^r f(p)(X_{\pi(1)}, \ldots, X_{\pi(r)}),$$

where $X_i \in V$ and π is a permutation of the numbers $1, \ldots, r$.

(b) If $f : U \to W$ and $g : U \to W$ are of class C^r on U, then so is $af + bg$, where a, $b \in R$ and $D^r(af + bg) = aD^r f + bD^r g$; that is, the C^r-functions form a vector space. Also if $W = R$, then the pointwise product fg is of class C^r; that is, $C^r(U)$ is an algebra over R.

(c) Let $f : U \to W$ be of class C^r. Then for all m, n with $m + n = r$ we have

$$D^{n+m} f = D^n D^m f$$

on U. That is, for $p \in U$,

$$D^{n+m} f(p)(X_1, \ldots, X_{n+m}) = [D^n[D^m f](p)(X_1, \ldots, X_n)](X_{n+1}, \ldots, X_{n+m}).$$

(d) Let $f : U \to R$ be a function and let $e_1, \ldots, e_n$ be the usual basis of $V = R^n$. Then $f \in C^r(U)$ if and only if all of the partial derivatives

$$\partial^q f(p)/\partial x_{i_1} \cdots \partial x_{i_q} = D_{i_1} \cdots D_{i_q} f(p)(e_{i_1}, \ldots, e_{i_q})$$

for $q = 1, \ldots, r$ exist for all $p \in U$ and are continuous on U. In this case we have for $X_1, \ldots, X_r \in V$,

$$D^r f(p)(X_1, \ldots, X_r) = \sum x_{1 i_1} \cdots x_{r i_r} \, \partial^r f(p)/\partial x_{i_1} \cdots \partial x_{i_r},$$

where the sum runs over all possible r-tuples $i_1, \ldots, i_r$ of $1, \ldots r$ and $X_j = \sum x_{jk} e_k$.

Exercises (2) Let $f : U \to W$, where $W = W_1 \times \cdots \times W_q$ and f is given by coordinate maps $(f_1, \ldots, f_q)$. Show f is of class C^r on U if and only if each $f_i : U \to W_i$ is of class C^r on U. In this case show $D^r f = (D^r f_1, \ldots, D^r f_q)$.

(3) Let U be open in V and E be open in W. Let $f : U \to E$ and $g : E \to Z$, where Z is some Euclidean space, both be C^r-functions. Then show the compositions $g \circ f : U \to Z$ is of class C^r on U. Also show $D^r(T \circ f)(p) = T \circ D^r f(p)$ for $T \in L(W, Z)$ and $p \in U$. In particular this holds for C^∞-functions.

(4) Let $f : U \to W$ be of class C^r on U, let $A_1, \ldots, A_{r-1} \in V$, and let

$$F : U \to W : x \to D^{r-1} f(x)(A_1, \ldots, A_{r-1}).$$

Show F is differentiable on U and

$$DF(p)(X) = D^r f(p)(A_1, \ldots, A_{r-1}, X)$$

for $X \in V$ and $p \in U$. This is the usual way for computing higher derivatives in terms of lower derivatives.

(5) (i) Show that the exponential mapping

$$\exp : \mathrm{End}(V) \to GL(V) : X \to e^X$$

is of class C^∞ on a suitable ball about 0 in the Euclidean space End(V). Also show

$$[D^2 \exp(0)](X, X) = X^2$$

and find a formula for the bilinear form $[D^2 \exp(0)](X_1, X_2)$. Show that $[D^r \exp(0)](X, \ldots, X) = X^r$.

(ii) More generally, let $\sum a_n t^n$ be a real power series with radius of convergence ρ. Then what can be said about the C^∞ nature of the function $f(T) = \sum a_n T^n$ in End(V)?

(6) Let n be a positive integer and let $X \in \mathrm{End}(V)$ be fixed. Define

$$f_n : GL(V) \to GL(V) : P \to P^{-n},$$
$$g_n : \mathrm{End}(V) \to \mathrm{End}(V) : P \to P^n,$$
$$h : \mathrm{End}(V) \to \mathrm{End}(V) : P \to PXP.$$

Show that:

(i) $[Dg_n(P)](X) = \sum_{k=0}^{n-1} P^k X P^{n-k-1}$,

(ii) $[Df_n(P)](X) = -\sum_{k=0}^{n-1} P^{k-n} X P^{-k-1}$,

(iii) $[Dh(P)]Y = PXY + YXP$,

(iv) $[D^2 f_1(P)](X, Y) = P^{-1}XP^{-1}YP^{-1} + P^{-1}YP^{-1}XP^{-1}$.

4. Taylor's Formula

We shall now discuss Taylor's formula for vector-valued functions which, when restricted to real-valued functions, gives the usual polynomial approximation of such functions. Also this is the formula which is used when the multiplication of a Lie group is approximated at the identity element by its Lie algebra. This is the approximation which allows the analysis of a Lie group by algebraic methods. The vector spaces in this section are still real Euclidean spaces.

Taylor's formula Let U be open in V and let $f : U \to W$ be of class C^r on U. Let $p \in U$ and $X \in V$ be such that the line segment $p + tX$ for $0 \leq t \leq 1$ is in U. Let $X^{(k)} = (X, \ldots, X)$ k-times and $D^k f(p)X^{(k)} = D^k f(p)(X, \ldots, X)$. Then

$$f(p + X) = f(p + Df(p)X + \frac{D^2 f(p)}{2!} X^{(2)} + \cdots + \frac{D^r f(p)}{r!} X^{(r)} + \varepsilon_r(X),$$

where the error term satisfies

$$\lim_{X \to 0} \varepsilon_r(X)/\|X\|^r = 0.$$

For the proof and other estimates for the error term we refer to Dieudonné [1960] and Lang [1968]. However, this formulation is the most practical for our use.

Definition 1.13 Let V and W be Euclidean spaces with U open in V. A function $f : U \to W$ is (real) **analytic on** U if f is of class C^∞ on U and if for each $p \in U$, there exists an open ball $B \subset U$ with center p so that for all $q = p + X$ in B the series

$$\sum_{r=0}^{\infty} \frac{1}{r!} D^r f(p) X^{(r)}$$

converges absolutely in the Euclidean space topology and has value $f(q)$. The function $f : U \to W$ is **analytic at** p **in** U if f is analytic on some neighborhood of p. We shall denote the set of functions $f : U \to W$ which are real analytic on U by $C^\omega(U, W)$ or $\mathscr{A}(U, W)$; see Dieudonné [1960] for more results on analytic functions.

Examples (1) Let $f : R \to R$ be defined by

$$f(x) = \begin{cases} e^{-1/x} & \text{for} \quad x > 0, \\ 0 & \text{for} \quad x \leq 0. \end{cases}$$

Then it is easy to see that $f \in C^\infty(R)$ and by induction that $D^k f(0) = 0$. Thus since f is not identically 0 in any neighborhood of 0, f is not analytic at 0. Otherwise in a suitable open interval U about 0,

$$f(x) = \sum D^k f(0) \frac{x^k}{k!} = 0$$

for all $x \in U$. Thus the classes of C^∞ and analytic functions are definitely different.

(2) Let V, W, and Z be Euclidean spaces and let U be an open neighborhood of 0 in V and D an open neighborhood of 0 in W. Let

$$f : U \to D \qquad \text{and} \qquad g : D \to Z$$

be functions of class C^3 so that $f(0) = 0$. Assuming f, g and $f \circ g$ satisfy the hypothesis of Taylor's formula we have the following second-order approximation for $X \in V$ at $p = 0$.

$$\begin{aligned}(g \circ f)(X) &= g(f(X)) \\ &= g(0) + Dg(0)(f(X)) + \frac{D^2g(0)}{2!}(f(X), f(X)) + \varepsilon_2 \\ &= g(0) + Dg(0)\left[Df(0)X + \frac{D^2f(0)}{2!}(X, X)\right] \\ &\quad + \frac{D^2g(0)}{2!}\left[Df(0)X + \frac{D^2f(0)}{2!}(X, X), Df(0)X + \frac{D^2f(0)}{2!}(X, X)\right] \\ &\quad + \varepsilon_2 \\ &= g(0) + Dg(0)Df(0)X + \tfrac{1}{2}Dg(0)[D^2f(0)(X, X)] \\ &\quad + \tfrac{1}{2}D^2g(0)\big(Df(0)X, Df(0)X\big) + \varepsilon_2,\end{aligned}$$

where

$$\lim_{X \to 0} \varepsilon_2(X)/\|X\|^2 = 0.$$

We shall use this later in an example on the matrix group $GL(V)$ in Section 1.6.

It is easy to see that if a function $f : U \to W$ is represented in terms of coordinates, then f is analytic on U if and only if each coordinate function is a (real-valued) analytic function on U. Thus if for each $x \in U$ and a basis $Y_1, \ldots, Y_m$ of W we have

$$f(x) = \sum_{j=1}^{m} f_j(x)Y_j,$$

then f is analytic if and only if for each $p \in U$ there exists an open ball $B \subset U$ with center p and m power series P_j, $j = 1, \ldots, m$, in n variables so that

$$f_j(q) = P_j(q_1 - p_1, \ldots, q_n - p_n)$$

for $q \in B$, where a basis $X_1, \ldots, X_n$ of V gives a coordinate system in V.

As an application of this formulation of analyticity we consider the Euclidean space $\mathrm{End}(V)$ relative to the sup norm and the exponential mapping

$$\exp : \mathrm{End}(V) \to GL(V) : X \to e^X.$$

From Section 1.3 we have for fixed a that exp is of class C^∞ on the open set $U = \{X \in \mathrm{End}(V) : \|X\| < a\}$ and since

$$[D^k \exp(0)]X^{(k)} = X^k$$

we have the rth order Taylor's formula approximation is

$$\exp X = I + X + \frac{1}{2} X^2 + \cdots + \frac{1}{r!} X^r + \varepsilon_r(X),$$

where

$$\varepsilon_r(X) = \sum_{k=r+1}^{\infty} X^k/k!.$$

We shall now show that the exponential series converges absolutely. First we note from example (1) of Section 1.1 by induction

$$\|X^k\| \leq \|X\|^k$$

and that the real series $\sum(\|X\|^k/k!)$ $(=e^{\|X\|})$ converges. Thus since

$$\sum\|X^k\|/k! \leq \sum\|X\|^k/k!$$

we have that the exponential series is absolutely convergent. If we restrict ourselves to a closed ball of radius a

$$\overline{B(0, a)} = \{X \in \mathrm{End}(V) : \|X\| \leq a\},$$

then since the series $\Sigma a^n/n!$ is convergent, we see from the Weierstrass test for series that the exponential series is uniformly convergent on compact subsets of End(V).

Now the exponential function is analytic on the open ball $B(0, a)$ which is contained in the compact ball of radius a. By choosing the usual basis of End(V) we determine coordinates $x_{11}, \ldots, x_{nn}$ and represent any $X \in U$ by the matrix (x_{ij}). Next we observe that each coordinate function for the partial sum of the series $\sum X^n/n!$ is a polynomial in $x_{11}, \ldots, x_{nn}$. Thus for each $p \in B(0, a)$ we can find a suitable ball B in $B(0, a)$ so that each of these polynomials is of the form $p_{ij}(q_{11} - p_{11}, \ldots, q_{nn} - p_{nn})$ on B. Therefore by the uniform convergence, each coordinate function for the series has a power series representation

$$P_{ij}(q_{11} - p_{11}, \ldots, q_{nn} - p_{nn})$$

about p.

Proposition 1.14 (a) Let V be a Euclidean space. Then the exponential mapping

$$\exp : \mathrm{End}(V) \to GL(V) : X \to e^X$$

is analytic on End(V).

(b) If $S, T \in \operatorname{End}(V)$ are such that $ST = TS$, then

$$\exp(T + S) = \exp T \exp S.$$

(c) For a fixed $S \in \operatorname{End}(V)$, the map

$$\phi : R \to GL(V) : t \to \exp tS$$

is an analytic homomorphism of the additive group of R into the multiplicative group $GL(V)$. In particular $(\exp S)^n = \exp(nS)$ for any integer n.

Proof First note that the vector space $\operatorname{End}(V) \times \operatorname{End}(V)$ is a Euclidean space and in terms of coordinates the multiplication function μ of endomorphisms is an analytic map

$$\mu : \operatorname{End}(V) \times \operatorname{End}(V) \to \operatorname{End}(V) : (S, T) \to S \circ T$$

where $S \circ T = \mu(S, T)$. Now the curves

$$f(t) = \exp(T + tS) \qquad \text{and} \qquad g(t) = \exp T \exp tS$$

are defined for all $t \in R$. These are analytic at $t = 0$ because the multiplication of endomorphisms is analytic and exp is analytic on any ball $B(0, a)$ as discussed above. Now we use some results from elementary differential equations (which we shall consider in Chapter 2) and note that since $ST = TS$ we have $(\exp S)T = T(\exp S)$. Consequently f and g are solutions to the differential equation

$$dY(t)/dt = SY(t) \quad \text{with} \quad Y(0) = \exp T.$$

However, by the uniqueness of solutions we have $f(t) = g(t)$ in an interval about 0. Since $f(t)$ and $g(t)$ are defined for all $t \in R$ and analytic functions of t at 0, they are equal for all $t \in R$. Thus for $t = 1$ we obtain (b).

For $s, t \in R$ note that the endomorphisms sS, tS commute and therefore from (b) we see

$$\phi(s + t) = \exp(s + t)S = \exp sS \exp tS = \phi(s)\phi(t).$$

We have shown that exp is analytic on any ball $B(0, a)$. Thus exp is analytic at any $p \in \operatorname{End}(V)$ by choosing the ball $B(0, a)$ above to contain p.

Exercises (1) (i) Let U be open in V and D be open in W and let

$$f : U \to W \qquad \text{and} \qquad g : D \to Z$$

be analytic functions so that $f(U) \subset D$. Then show the composition $g \circ f$ is analytic on U.

(ii) Let U be open in V and suppose $f, g \in \mathscr{A}(U, R)$, the set of real-valued analytic functions on U. Show that $\mathscr{A}(U, R)$ is an algebra of functions relative to the pointwise product.

(2) Show that the Taylor's series representation for an analytic function is unique. That is, if $f : U \to W$ is analytic and if for $p \in U$ and $X \in V$ satisfying the analyticity conditions we have for all t with $0 \leq t \leq 1$ that

$$f(p + tX) = \sum_{k=0}^{\infty} a_k(p)(tX)^{(k)} = \sum_{k=0}^{\infty} b_k(p)(tX)^{(k)}$$

with absolute convergence, then show $a_k(p) = b_k(p)$ as multilinear forms.

(3) (i) In more detail, why is the map

$$\phi : R \to GL(V) : t \to \exp tS$$

used in Proposition 1.14 analytic?

(ii) Show that the functions f and g used in the proof of Proposition 1.14 actually satisfy the indicated differential equation.

(4) Let $\lambda \in L^n(V, W)$ be a multilinear map from V^n to W and let $P = (P_1, \ldots, P_n)$ and $X = (X_1, \ldots, X_n)$ be in V^n.

(i) Show $[D\lambda(P)](X) = \lambda(X_1, P_2, \ldots, P_n) + \cdots + \lambda(P_1, P_2, \ldots, P_{n-1}, X_n)$.

(ii) Find $[D^n\lambda(P)]X^{(n)}$.

(iii) Find the Taylor's series expansion for λ about $P = (0, \ldots, 0)$.

(5) For a fixed basis of $V = R^n$ let $A \in \mathrm{End}(V)$ be represented by the matrix $\alpha = (a_{ij})$ so that α can be considered as an element of $R^n \times \cdots \times R^n$ by viewing the rows of α as elements of R^n. Consequently the determinant function $\det : \mathrm{End}(V) \to R$ can be considered in terms of the coordinates given by the above matrix representation.

(i) Show $\det : \mathrm{End}(V) \to R$ is differentiable (of what class?).

(ii) In terms of matrices, let $a_1, \ldots, a_n \in R^n$ denote the rows of α and $x_1, \ldots, x_n$ denote the rows of any $n \times n$ matrix X. Show

$$[D \det(\alpha)]X = \sum_{i=1}^{n} \det \begin{bmatrix} a_1 \\ \vdots \\ x_i \\ \vdots \\ a_n \end{bmatrix}$$

(6) There exist differentiable "bump functions" which vanish outside a compact set as follows. Show if $B(0, a)$ and $B(0, b)$ with $a < b$ are two

concentric balls about $0 \in R^m$, then there exists a C^∞-function $\phi : R^m \to R$ such that $0 \le \phi(x) \le 1$ for all $x \in R^m$ and

$$\phi(q) = \begin{cases} 1 & \text{if} \quad q \in B(0, a), \\ 0 & \text{if} \quad q \notin B(0, b). \end{cases}$$

(*Hint*: Use the function f of example (1), Section 1.4 and consider $\phi(x) = f(b^2 - \|x\|^2)/[f(b^2 - \|x\|^2) + f(\|x\|^2 - a^2)]$).

5. Inverse Function Theorem

We continue the notation of the preceding sections and investigate when a function $f: U \to W$ has a local inverse and the various consequences. The vector spaces are still real Euclidean spaces.

Definition 1.15 Let $f: U \to W$ be of class C^r on U and let $p \in U$. Then f is **locally invertible of class C^r at p** if there exists an open subset U_1 of U with $p \in U_1$ such that there exists an open set D of W with $f(p) \in D$ and a function $g : D \to U_1$ of class C^r on D such that $g \circ f$ and $f \circ g$ are the identity maps on U_1 and D, respectively. The function g is called a **local inverse** of f at $f(p)$.

With this notation we have the following (see Dieudonné [1960], Lang [1968], and Spivak [1965]).

Inverse Function Theorem Let U be open in V, let $p \in U$, and let $f : U \to W$ be of class C^r on U. If $Df(p) \in \operatorname{Hom}(V, W)$ is an invertible linear transformation, then f is locally invertible of class C^r at p. Furthermore if g is a local inverse of f and $q = f(p)$, then $Dg(q) = [Df(p)]^{-1}$.

REMARKS (1) Since the vector spaces V and W are finite-dimensional, $Df(p)$ being invertible yields $\dim V = \dim W$.

(2) Let $f : U \to D$ be continuous with a continuous inverse $g : D \to U$ and let U_1 be an open subset of U. Then $f(U_1)$ is an open subset of D. Thus in particular if $f : U \to W$ is of class C^1 and $Df(p)$ is invertible, then for a suitable open subset U_1 of U with $p \in U_1$, we have $f(U_1)$ is open in W and $f: U_1 \to f(U_1)$ is a homeomorphism.

(3) A C^∞-function can have a continuous inverse without the inverse being differentiable. Thus the function $f : R \to R : x \to x^3$ is of class C^∞ on R and has continuous inverse $g : R \to R : y \to y^{1/3}$. However, g is not differentiable at the point 0 (why?).

(4) The various results concerning the inverse function theorem also apply to C^∞ and analytic functions.

The inverse function theorem is used to obtain the following result.

Theorem 1.16 (a) Let V_1, V_2, W be real Euclidean vector spaces, let U be open in $V_1 \times V_2$ and let $f : U \to W$ be of class C^r on U. If for $(a, b) \in U \subset V_1 \times V_2$ the linear transformation $D_2 f(a, b) : V_2 \to W$ is invertible, then $\dim V_2 = \dim W$ and the map

$$F : U \to V_1 \times W : (x, y) \to (x, f(x, y))$$

is locally invertible of class C^r at (a, b).

(b) (Implicit function theorem) Let U be open in $V_1 \times V_2$ and let $f : U \to W$ be a map of class C^r such that for a given $(a, b) \in U, f(a, b) = 0$. If $D_2 f(a, b) : V_2 \to W$ is invertible, then there exists an open ball B in V_1 with center at $a \in V_1$, and there exists a uniquely determined map $g : B \to V_2$ of class C^r on B such that

$$g(a) = b \qquad \text{and} \qquad f(x, g(x)) = 0$$

for all $x \in B$.

PROOF Briefly for part (a) we note that since $D_2 f(a, b)$ is invertible, $\dim V_2 = \dim W$. Thus writing F in coordinates, we have $F(x, y) = (F_1(x, y), F_2(x, y))$ where $F_1(x, y) = x$ and $F_2(x, y) = f(x, y)$. Consequently $DF(a, b)$ has a matrix given by

$$\begin{bmatrix} D_1 F_1(a, b) & D_2 F_1(a, b) \\ D_1 F_2(a, b) & D_2 F_2(a, b) \end{bmatrix} = \begin{bmatrix} I & 0 \\ D_1 f(a, b) & D_2 f(a, b) \end{bmatrix}$$

which is invertible. Now apply the inverse function theorem.

For part (b) let

$$h : U \to V_1 \times W : (x, y) \to (x, f(x, y)).$$

Then by part (a), h has a local inverse denoted by H. Thus since H is a mapping into $V_1 \times V_2$, it can be represented by coordinate functions $H = (H_1, H_2)$ such that

$$H(x, z) = (H_1(x, z), H_2(x, z)) \equiv (x, H_2(x, z)).$$

Then the desired function is given by $g(x) = H_2(x, 0)$. To see that $y = g(x)$ satisfies the equation, we have for $z = 0$ that

$$\begin{aligned} (x, 0) &= (h \circ H)(x, 0) \\ &= h(x, H_2(x, 0)) \\ &= h(x, g(x)) = (x, f(x, g(x))). \end{aligned}$$

To see that $g(a) = b$, we note

$$\begin{aligned}(a, g(a)) &= (a, H_2(a, 0)) \\ &= (a, H_2(a, f(a, b))) \\ &= (H \circ h)(a, b) = (a, b).\end{aligned}$$

The rest of the proof is not difficult and can be found in the work of Lang [1968] and Spivak [1965].

Example (1) The surface given by $z = f(x, y) = x^2 + y^2 - 1$ defines a circle for those points (x, y) with $f(x, y) = 0$. Choose (a, b) on the circle with $b \neq 0$. Then $\partial f(a, b)/\partial y = 2b \neq 0$ so that we can solve locally for y in terms of x.

A variation of the preceding proof gives the following result which we shall use in Chapter 2.

Proposition 1.17 Let U be an open set in R^n which contains the point p and let $f : U \to R^q$ be of class C^r on U where $q \leq n$. If $f(p) = 0$ and $Df(p)$ has rank q, then there exist an open subset D of R^n containing p and a locally invertible function $g : D \to U$ of class C^r such that for all $(x_1, \ldots, x_n) \in D$,

$$(f \circ g)(x_1, \ldots, x_n) = (x_{n-q+1}, \ldots, x_n).$$

Thus one can modify the coordinates of the point p by the open set D and the function g to obtain a simple expression for f locally.

PROOF We regard $R^n = R^{n-q} \times R^q$ and U as an open set in $R^{n-q} \times R^q$. Thus $p = (a, b) \in U \subset R^{n-q} \times R^q$ is such that regarding f as a function of two variables we have $f(a, b) = f(p) = 0$. Since $Df(p)$ is of rank q we see $D_2 f(a, b)$ is of rank q (why?); that is, $D_2 f(a, b) : R^q \to R^q$ is invertible. Thus we are in the situation of the preceding theorem as follows. If we let $x = (x_1, \ldots, x_{n-q})$, $z = (x_{n-q+1}, \ldots, x_n)$, and $g(x_1, \ldots, x_n) = H(x, z)$ as in the above proof, we have for $\pi : (x, z) \to z$ that

$$\begin{aligned}(f \circ g)(x, z) &= (f \circ H)(x, z) \\ &= (\pi \circ h \circ H)(x, z) \\ &= (\pi \circ idy)(x, z) = z.\end{aligned}$$

This also uses the notation of the preceding proof $h(x, y) = (x, f(x, y))$ so that $f = \pi \circ h$, where H is the local inverse of h.

Definition 1.18 Let U_i be open subsets of the Euclidean spaces V_i for $i = 1, 2$. Then the map f: $U_1 \to U_2$ is a **diffeomorphism of class C^r** of U_1 onto U_2 if f is a homeomorphism of U_1 onto U_2 so that f is of class C^r on U_1 and

f^{-1} is of class C^r on U_2. We analogously define **analytic diffeomorphisms** and say f is a **diffeomorphism** if f is a diffeomorphism of class C^∞.

We continue our previous notation for the exponential mapping and have the following result.

Proposition 1.19 There exists an open neighborhood U_0 of 0 in $\mathrm{End}(V)$ and an open neighborhood U_I of I in $GL(V)$ such that $\exp : U_0 \to U_I$ is a diffeomorphism of U_0 onto U_I. Furthermore exp and its inverse are actually analytic.

PROOF From the inverse function theorem and previous results, it suffices to show that $D\exp(0)$ is invertible. Thus using the series expansion, we have for $A \in \mathrm{End}(V)$ that

$$\begin{aligned}[D\exp(0)](A) &= \lim_{t\to 0} \frac{1}{t}\,[\exp(0 + tA) - \exp(0)] \\ &= \lim_{t\to 0} \frac{1}{t}\,[I + tA + \varepsilon(t^2) - I] = A\end{aligned}$$

using $\lim_{t\to 0} \varepsilon(t^2)/t = 0$. Thus $D\exp(0)$ is the identity.

REMARKS (5) For any $X \in \mathrm{End}(V)$ we have $tX \in U_0$ for $t \in R$ in a suitable interval N of 0 in R. Thus exp maps this line segment tX with $t \in N$ into an analytic curve segment $\phi(t) = \exp tX$ in $GL(V)$ which passes through I; that is, $\exp(0) = I$ and the "tangent vector" to this curve at $t = 0$ is X.

(6) By considering the power series

$$\log(1 + z) = z - \frac{z^2}{2} + \cdots + (-1)^{n+1}\frac{z^n}{n} + \cdots$$

and the proposition in Section 1.1 we have

$$\log(I + A) = A - \frac{A^2}{2} + \cdots + (-1)^{n+1}\frac{A^n}{n} + \cdots$$

converges when the absolute value of the characteristic roots of A is less than 1. Now if we use the usual sup norm for $T \in \mathrm{End}(V)$, we have

$$\begin{aligned}\|T\| &= \sup_{X\neq 0} \|TX\|/\|X\| \\ &\geq \max\{|r_1|, \ldots, |r_n|\},\end{aligned}$$

where $r_1, \ldots, r_n$ are the characteristic roots of T and $\| \ \|$ on the right side of the first equation denotes length in V. Thus choosing $U_0 \subset \{A \in \mathrm{Hom}(V, V) : \|A\| < 1\}$ and $U_I \subset \{B \in \mathrm{Hom}(V, V) : \|B - I\| < 1\}$, we can show

$$\exp : U_0 \to U_I \qquad \text{and} \qquad \log : U_I \to U_0$$

and

$$\log \exp A = A \qquad \text{and} \qquad \exp \log B = B.$$

Thus log is the local inverse of exp and this inverse is analytic on U_I.

Exercises (1) Find two different 2×2 real matrices A and B such that $\exp A = \exp B$; that is, exp is not globally injective.

(2) Use Taylor's formula or the mean value theorem to prove a variation of the inverse function theorem.

6. The Algebra $gl(V)$

We shall now use many of the preceding results on the function exp to attach a nonassociative algebra $gl(V)$ to the general linear group $GL(V)$. The relationship between the group $GL(V)$ and the algebra $gl(V)$ is the basic model for studying a Lie group by its Lie algebra and we break down the analysis as follows.

(1) The map exp is used to show that the global multiplication in $GL(V)$ induces a local analytic multiplication in $\mathrm{End}(V)$.

(2) We use the second derivative of the local multiplication in $\mathrm{End}(V)$ to obtain a bilinear multiplication τ on $\mathrm{End}(V)$ which together yield the algebra $gl(V)$.

(3) The properties of $GL(V)$ are used to obtain a formula for τ and the identities it satisfies.

(4) Finally we show how automorphisms of $GL(V)$ induce automorphisms of $gl(V)$ and indicate some important formulas.

For the first step let

$$\mu : GL(V) \times GL(V) \to GL(V) : (x, y) \to \mu(x, y) \equiv xy$$

be the multiplication in $GL(V)$. Then since matrix multiplication is given in terms of polynomials of degree 2, the multiplication μ is analytic. In particular by the continuity of μ we have for any neighborhood U of $I \in GL(V)$ that there exists a neighborhood D of I so that

$$D \subset U \qquad \text{and} \qquad \mu(D, D) \subset U.$$

However, using the continuity of the map $\exp : \mathrm{End}(V) \to GL(V)$ we can find for any neighborhood D of $I \in GL(V)$ a neighborhood E of $0 \in \mathrm{End}(V)$ so that

$$\exp(E) \subset D.$$

Now we choose the above neighborhoods according to Proposition 1.19 as follows: Let $U \subset U_I$ so that $E \subset U_0$ and $\exp(E) \subset D \subset U_I$. Thus

$$\mu(\exp(E), \exp(E)) \subset U_I.$$

This means, using $\exp U_0 \supset U_I$, that we can define a function analytic on $E \times E$

$$F : E \times E \to U_0 : (X, Y) \to F(X, Y)$$

so that for $X, Y \in E$,

$$\mu(\exp X, \exp Y) = \exp F(X, Y).$$

From the remarks following Proposition 1.19 we see

$$F(X, Y) = \log \mu(\exp X, \exp Y).$$

Thus the multiplication μ in $GL(V)$ induces a local analytic multiplication F in $\mathrm{End}(V)$; that is, we have neighborhoods U_0 and E of 0 in $\mathrm{End}(V)$ so that $E \subset U_0$ and a "multiplication" $F : E \times E \to U_0$.

A variation of the above is as follows. Let X and Y be any (fixed) elements of $\mathrm{End}(V)$. Then there exists a neighborhood N of 0 in R so that for all $s, t \in N$ we have $sX, tY \in E$. Consequently we have the formula

$$\mu(\exp sX, \exp tY) = \exp F(sX, tY)$$

which gives an analytic function

$$N \times N \to U_0 : (s, t) \to F(sX, tY).$$

Thus since F is analytic on $E \times E$ we have the Taylor's series expansion about $\theta = (0, 0) \in E \times E$ as follows. Let $Z = (sX, tY)$ be as above. Then $\theta + Z = (sX, tY) = Z$ so that

$$F(Z) = F(\theta) + DF(\theta)Z + \frac{D^2F(\theta)}{2!} Z^{(2)} + \cdots.$$

In particular for $s = t$ we obtain

$$F(sX, sY) = F(\theta) + sDF(\theta)(X, Y) + s^2 \frac{D^2F(\theta)}{2!}((X, Y), (X, Y)) + \cdots.$$

For the second step we now compute the first few terms of this Taylor's series. From $\exp 0 = I$ we obtain

$$\begin{aligned}\exp F(0,0) &= \mu(\exp 0, \exp 0)\\ &= \mu(I, I) = I = \exp 0\end{aligned}$$

and since exp is injective on U_0 (that is, applying log) we have

$$F(\theta) = F(0,0) = 0.$$

Similarly since $\exp 0$ is the identity element in $GL(V)$ we have

$$\exp sX = \mu(\exp sX, \exp 0) = \exp F(sX, 0)$$

so that

$$\begin{aligned}sX &= F(sX, 0)\\ &= sDF(\theta)(X, 0) + \sum_{k=2}^{\infty} s^k \frac{D^kF(\theta)}{k!}(X, 0)^{(k)}.\end{aligned}$$

Thus by comparing degrees (or differentiating term-by-term at $s = 0$) we obtain

$$DF(\theta)(X, 0) = X \qquad \text{and} \qquad D^kF(\theta)(X, 0)^{(k)} = 0 \qquad \text{for} \quad k \geq 2.$$

Also we obtain

$$DF(\theta)(0, Y) = Y \qquad \text{and} \qquad D^kF(\theta)(0, Y)^{(k)} = 0 \qquad \text{for} \quad k \geq 2$$

so that we have the following two formulas.

$$\begin{aligned}DF(\theta)Z &= DF(\theta)(sX, tY)\\ &= DF(\theta)[(sX, 0) + (0, tY)]\\ &= sDF(\theta)(X, 0) + tDF(\theta)(0, Y) = sX + tY\end{aligned}$$

and

$$\begin{aligned}D^2F(\theta)Z^{(2)} &= D^2F(\theta)[(sX, 0) + (0, tY), (sX, 0) + (0, tY)]\\ &= D^2F(\theta)(sX, 0)^{(2)} + D^2F(\theta)(0, tY)^{(2)} + 2D^2F(\theta)[(sX, 0), (0, tY)]\\ &= 2stD^2F(\theta)[(X, 0), (0, Y)],\end{aligned}$$

where the last two equalities use the symmetric bilinearity of $D^2F(\theta)$ on $\mathrm{End}(V) \times \mathrm{End}(V)$ and $D^2F(\theta)(X, 0)^{(2)} = D^2F(\theta)(0, Y)^{(2)} = 0$.

From $D^kF(\theta)(X, 0)^{(k)} = D^kF(\theta)(0, Y)^{(k)} = 0$ we see that in the Taylor's series expansion for $F(sX, tY)$ the coefficients of s^k and t^k are 0. Consequently the error term

$$\varepsilon_2(sX, tY) = st\phi(s, t),$$

where $\phi(s, t) \to 0$ as $s \to 0$ and $t \to 0$. Thus Taylor's formula for the local multiplication becomes

$$F(sX, tY) = sX + tY + st\tau(X, Y) + st\phi(s, t),$$

where

$$\tau(X, Y) = D^2F(\theta)[(X, 0), (0, Y)] = \tfrac{1}{2}D^2F(\theta)(X, Y)^{(2)}.$$

Next, since the second derivative is a quadratic function defined on the whole vector space, we see $\tau(X, Y)$ is defined for all $X, Y \in \mathrm{End}(V)$. Also

$$\tau : \mathrm{End}(V) \times \mathrm{End}(V) \to \mathrm{End}(V)$$

is bilinear as follows. For $X, Y, U \in \mathrm{End}(V)$

$$\begin{aligned}\tau(X + Y, U) &= D^2F(\theta)[(X + Y, 0), (0, U)]\\ &= D^2F(\theta)[(X, 0), (0, U)] + D^2F(\theta)[(Y, 0), (0, U)]\\ &= \tau(X, U) + \tau(Y, U).\end{aligned}$$

Similarly $\tau(U, X + Y) = \tau(U, X) + \tau(U, Y)$ and for $a \in R$, $\tau(aX, Y) = \tau(X, aY) = a\tau(X, Y)$. Thus the vector space $\mathrm{End}(V)$ together with the bilinear map $\tau : \mathrm{End}(V) \times \mathrm{End}(V) \to \mathrm{End}(V)$ becomes a nonassociative algebra as follows.

Definition 1.20 (a) A **nonassociative algebra** A over a field K is a vector space A together with a bilinear multiplication $\tau \in L^2(A, A)$; note that "nonassociative" means not necessarily associative.

(b) The nonassociative algebra $gl(V)$ with multiplication τ as above is called the **Lie algebra** of $GL(V)$.

From the above formulas we see that the algebra $gl(V)$ determines the multiplication in $GL(V)$ locally up to order two. Thus for s, t in a suitable neighborhood of 0 in R we have for all $X, Y \in gl(V)$

$$\mu(\exp sX, \exp tY) = \exp\big(sX + tY + st\tau(X, Y) + \varepsilon_2(sX, tY)\big), \qquad (*)$$

where $X + Y$ is the addition in $gl(V)$ and $\tau(X, Y)$ the multiplication. We shall show later (Campbell–Hausdorff theorem) that the error term $\varepsilon_2(X, Y)$ is actually determined by the subalgebra of $gl(V)$ generated by X and Y.

For the third step we use various properties of $GL(V)$ to determine identities for $gl(V)$.

Proposition 1.21 If $X, Y, Z \in gl(V)$, then

(a) $\tau(X, Y) = -\tau(Y, X)$;
(b) $\tau(X, \tau(Y, Z)) + \tau(Y, \tau(Z, X)) + \tau(Z, \tau(X, Y)) = 0$;
(c) $\tau(X, Y) = \tfrac{1}{2}(XY - YX)$.

PROOF (a) Since $(\exp X)^{-1} = \exp(-X)$ the inverse of the left-hand side of $(*)$ above becomes

$$\exp(-tY)\exp(-sX) = \exp\bigl(-tY - sX + st\tau(Y, X) + \cdots\bigr)$$

and the inverse of the right-hand side of $(*)$ becomes

$$\exp(-sX - tY - st\tau(X, Y) + \cdots).$$

The required identity now follows from comparing the terms of degree 2.

(c) The two sides of $(*)$ can be expanded by power series methods using the multiplication in $\mathrm{End}(V)$ as follows

$$\begin{aligned}\exp(sX)\exp(tY) &= (I + sX + \tfrac{1}{2}s^2X^2 + \cdots)(I + tY + \tfrac{1}{2}t^2Y^2 + \cdots)\\ &= I + sX + tY + \tfrac{1}{2}s^2X^2 + stXY + \tfrac{1}{2}t^2Y^2 + \cdots,\\ \exp\bigl(sX + tY + st\tau(X, Y) + \cdots\bigr) &= I + \bigl(sX + tY + st\tau(X,Y) + \cdots\bigr)\\ &\quad + \tfrac{1}{2}(sX + tY + \cdots)^2 + \cdots\\ &= I + sX + tY + st\tau(X, Y)\\ &\quad + \tfrac{1}{2}(s^2X^2 + stXY + stYX + t^2Y^2) + \cdots,\end{aligned}$$

where the omitted terms are of degree greater than or equal to 3 in s and t. Equating the coefficients of st in the two expressions we have

$$XY = \tau(X, Y) + \tfrac{1}{2}XY + \tfrac{1}{2}YX$$

and the desired identity now follows.

(b) This can be computed directly by substituting (c) into the left side of (b) and using the associativity of multiplication in $\mathrm{End}(V)$.

REMARKS (1) Part (c) above could have been proved directly as indicated and the property (a) obtained from this. We shall also indicate how to obtain (b) from facts on automorphisms and Taylor's series.

(2) An abstract **Lie algebra** over a field K is defined to be a nonassociative algebra whose multiplication satisfies the **anticommutativity** and **Jacobi identity** given, respectively, by (a) and (b) above.

(3) When just the Lie algebra $gl(V)$ is considered the multiplication is usually denoted by

$$[X, Y] = XY - YX$$

which is called the **commutator** of X and Y.

For the fourth part on automorphisms we manipulate more Taylor's series.

Proposition 1.22 Let f be an analytic diffeomorphism of $GL(V)$ which is also an automorphism of the group $GL(V)$. Then $Df(I)$ is an automorphism of the Lie algebra $gl(V)$; that is, $Df(I) \in GL(\text{End}(V))$ and

$$Df(I)\tau(X, Y) = \tau(Df(I)X, Df(I)Y)$$

for all $X, Y \in \text{End}(V)$.

PROOF Since f is a differentiable automorphism with differentiable inverse f^{-1} we have $f^{-1} \circ f = idy$, the identity map on End(V). Therefore by the chain rule

$$\begin{aligned} idy &= D(idy)(I) \\ &= D(f^{-1} \circ f)(I) = Df^{-1}(f(I)) \circ Df(I) \end{aligned}$$

so that $Df(I)$ is invertible; that is, $Df(I) \in GL(\text{End}(V))$.

Next notice that from the continuity of f there exists a neighborhood D_0 of $0 \in \text{End}(V)$ so that $D_0 \subset U_0$ and $X \in D_0$ implies $f(\exp X) \in U_I$. Define

$$k : D_0 \to U_0 : X \to \log f(\exp X)$$

so that for $X \in D_0$,

$$f(\exp X) = \exp k(X).$$

From this last equation it easily follows that $k(0) = 0$ and that

$$\begin{aligned} (Df)(I) &= (Df)(\exp 0) \circ (D \exp)(0) \\ &= D(f \circ \exp)(0) \\ &= D(\exp \circ k)(0) \\ &= (D \exp)(k(0)) \circ (Dk)(0) = Dk(0) \end{aligned}$$

since $(D \exp)(0)$ is the identity. Thus

$$k(0) = 0 \qquad \text{and} \qquad Df(I) = Dk(0).$$

As earlier in this section define $F(X, Y) \in \text{End}(V)$ for X, Y in End(V) sufficiently close to 0 so that

$$\mu(\exp X, \exp Y) = \exp F(X, Y).$$

Then for X, Y close enough to 0 we have

$$\begin{aligned} \exp k(F(X, Y)) &= f(\exp F(X, Y)) \\ &= f(\mu(\exp X, \exp Y)) \\ &= \mu(f(\exp X), f(\exp Y)) \\ &= \mu(\exp k(X), \exp k(X)) \\ &= \exp F(k(X), k(Y)) \end{aligned}$$

so that

$$k(F(X, Y)) = F(k(X), k(Y)).$$

Let $\theta = (0,0)$ and $Z = (X, Y)$. Then the Taylor's series for a composite function calculated earlier yeilds

$$\begin{aligned}(k \circ F)(Z) &= Dk(0) \circ DF(\theta)Z + \tfrac{1}{2}Dk(0) \circ D^2F(\theta)Z^{(2)} \\ &\quad + \tfrac{1}{2}D^2k(0)\big(DF(\theta)Z\big)^{(2)} + \varepsilon_2(Z) \\ &= Dk(0)(X + Y) + Dk(0)\big(\tau(X, Y)\big) \\ &\quad + \tfrac{1}{2}D^2k(0)(X + Y)^{(2)} + \varepsilon_2(Z)\end{aligned}$$

since $DF(\theta)(Z) = X + Y$ and $D^2F(\theta)Z^{(2)} = 2\tau(X, Y)$. Similarly $F\big(k(X), k(Y)\big)$ has a Taylor's series

$$\begin{aligned}F\big(k(X), k(Y)\big) &= DF(\theta)\big(k(X), k(Y)\big) + \tfrac{1}{2}D^2F(\theta)\big(k(X), k(Y)\big)^{(2)} + \cdots \\ &= k(X) + k(Y) + \tau\big(k(X), k(Y)\big) + \cdots \\ &= Dk(0)(X) + Dk(0)(Y) + \tfrac{1}{2}D^2k(0)X^{(2)} \\ &\quad + \tfrac{1}{2}D^2k(0)Y^{(2)} + \tau\big(Dk(0)(X), Dk(0)(Y)\big) + \varepsilon_2'(Z),\end{aligned}$$

where $\varepsilon_2(Z)$ and $\varepsilon_2'(Z)$ in the two expansions involve only terms of degree greater than or equal to 3.

Now if we substitute sX for X and tY for Y in the two series, it is clear that the two coefficients of st must be equal. Thus

$$Dk(0)\tau(X, Y) + D^2k(0)(X, Y) = \tau\big(Dk(0)X, Dk(0)Y\big).$$

However, $D^2k(0)(X, Y)$ is symmetric in X and Y and τ is an antisymmetric bilinear map so clearly $D^2k(0)(X, Y) = 0$ and

$$Dk(0)\tau(X, Y) = \tau\big(Dk(0)X, Dk(0)Y\big).$$

Since $Dk(0) = Df(I)$ we have completed our proof.

Examples (1) Let X^t denote the transpose of $X \in \mathrm{End}(V)$ relative to the usual inner product in $V = R^n$; that is, $B(X^tP, Q) = B(P, XQ)$. Then the map

$$f: GL(V) \to GL(V) : X \to (X^t)^{-1}$$

is an automorphism of class C^∞ with differentiable inverse. Thus $Df(I)$ is an automorphism of $gl(V)$ and is given by

$$\begin{aligned}Df(I)Y &= \lim_{s \to 0} \frac{1}{s}\,[f(I + sY) - f(I)] \\ &= \lim_{s \to 0} \frac{1}{s}\,[(I^t + sY^t)^{-1} - I] \\ &= \lim_{s \to 0} \frac{1}{s}\,[I - sY^t + s^2(Y^t)^2 + \cdots - I] = -Y^t.\end{aligned}$$

(2) Next we consider inner automorphisms

$$f: GL(V) \to GL(V) : Y \to ZYZ^{-1}$$

for fixed $Z \in GL(V)$. Thus f is of class C^∞ with differentiable inverse and $Df(I)$ is an automorphism of $gl(V)$ given by

$$Df(I)Y = \lim_{s \to 0} \frac{1}{s} [Z(I + sY)Z^{-1} - Z(I)Z^{-1}] = ZYZ^{-1}.$$

Moreover if $Z = \exp U$ for suitable $U \in \mathrm{End}(V)$, then by multiplying power series we obtain

$$\begin{aligned} Df(I)Y &= (\exp U)Y(\exp - U) \\ &= Y + UY - YU + \tfrac{1}{2}U^2Y - UYU + \tfrac{1}{2}YU^2 + \cdots \\ &= Y + [U, Y] + \tfrac{1}{2}[U, [U, Y]] + \cdots. \end{aligned}$$

We shall use the notation

$$\mathrm{ad}\, U : gl(V) \to gl(V) : Y \to [U, Y]$$

and observing that

$$e^{\mathrm{ad}\, U}Y = [I + \mathrm{ad}\, U + \tfrac{1}{2}(\mathrm{ad}\, U)^2 + \cdots]Y = Y + [U, Y] + \tfrac{1}{2}[U, [U, Y]] + \cdots$$

we indicate below how to show

$$Df(I) = e^{\mathrm{ad}\, U}.$$

Exercises (1) Prove the Jacobi identity [Proposition 1.21(b)] for $gl(V)$ using inner automorphisms of $GL(V)$ possibly as follows:

(i) For $2\tau(X, Y) = [X, Y]$ the above formula for $Z = \exp tU$ becomes $Df(I)Y = Y + 2t\tau(U, Y) + \varepsilon(t^2)$, where $\varepsilon(t^2)/t \to 0$ as $t \to 0$. Apply this replacing Y by $\tau(X, Y)$.

(ii) Similarly compute $\tau\big(Df(I)X, Df(I)Y\big)$ and use part (i) with the fact that $Df(I)$ is an automorphism of the multiplication τ.

(2) Let $Z = e^U$ and let $f \equiv f(U) : Y \to ZYZ^{-1}$ as in example (2) above. To show $Df(I) = e^{\mathrm{ad}\, U}$ one can proceed as follows. Let $\alpha(t) = e^{t\, \mathrm{ad}\, U}\, Y$ for $t \in R$ and for $Y \in GL(V)$. Then note that $\alpha(0) = Y$ and $\alpha'(t) = \mathrm{ad}\, U\big(\alpha(t)\big)$. Similarly letting $\beta(t) = D[f(tU)](I)Y$, note that $\beta(0) = Y$ and $\beta(t) = e^{tU}Ye^{-tU}$. Consequently $\beta'(t) = \mathrm{ad}\, U\big(\beta(t)\big)$ and using the uniqueness results for differential equations we obtain $Df(I) = e^{\mathrm{ad}\, U}$.

CHAPTER 2

MANIFOLDS

A manifold is a topological space where some neighborhood of a point looks like an open set in a Euclidean space. Thus we are able to translate the calculus of the preceding chapter to this type of a space and we develop the formalism in this chapter. In the first few sections we consider differentiable structures, the definition of a manifold, and real-valued differentiable functions defined on a manifold. Next we consider submanifolds and how they arise from the inverse function theorem; we give many examples of submanifolds which are subgroups of $GL(n, R)$. The derivative is generalized to a tangent at a point p in a manifold M and then the vector space spanned by these tangents generalizes the tangent plane of a surface. As the point p varies over M we obtain a variable tangent vector which is formalized via vector fields. We give many examples concerning $GL(n, R)$ which will be abstracted in later chapters; in particular we consider the invariant vector fields on $GL(n, R)$ and their integral curves.

Most of this chapter is used in the rest of the book and the reader who knows this material need only look at the examples. However, if one is unfamiliar with manifolds it might be best to read through Section 2.3, then read Chapters 3 and 4 for applications before finishing this chapter. The reader should note that we are assuming a neighborhood of a point is an open set in the space.

1. Differentiable Structures

We now extend the basic concepts of Euclidean space to a topological space which locally looks like Euclidean space via suitable choices of "coordinates."

Definition 2.1 (a) Let $V = R^m$ and let $X_1, \ldots, X_m$ be a basis of V so that we can represent any point $p = \sum p_k X_k \in V$ uniquely. Relative to this basis, we define **coordinate functions** u_i for $i = 1, \ldots, m$ on R^m by

$$u_i : R^m \to R : \sum p_k X_k \to p_i .$$

We shall frequently use the usual orthonormal basis $e_1, \ldots, e_m$ to obtain the **usual coordinates** u_i given by $u_i(a_1, \ldots, a_m) = a_i$.

(b) Let M be a topological space and let $p \in M$. An m**-dimensional chart at** $p \in M$ is a pair (U, x), where U is an open neighborhood of p and x is a homeomorphism of U onto an open set in R^m. The **coordinates** of the chart (U, x) are the functions x_i for $i = 1, \ldots, m$ given by

$$x_i = u_i \circ x : U \to R : q \to x_i(q),$$

where $x_i(q) = u_i(x(q))$ and the u_i are coordinates in R^m. We frequently write $x = (x_1, \ldots, x_m)$. The set U is called a **coordinate neighborhood** and (U, x) is called a **coordinate system** at $p \in M$.

Definition 2.2 An m**-dimensional topological manifold** M is a Hausdorff space with a countable basis such that for every $p \in M$ there exists an m-dimensional chart at p. In this case we say that the **dimension of** M is m.

Thus in particular we can find a covering of M by open sets and each open set U in the covering is homeomorphic to the open m-ball $B_m = \{a \in R^m : \|a\| < 1\}$.

Examples (1) Any open subset N of R^m is a manifold of dimension m, since N itself is a coordinate neighborhood of each of its points and the identity map x is such that (N, x) is an m-dimensional chart. Thus for $V = R^n$ we have $GL(V) \subset R^{n^2}$ is a manifold of dimension n^2. Note that for a fixed basis in V, any linear transformation $A \in GL(V)$ has a unique matrix representation (a_{ij}) and coordinate functions x_{ij} can be defined by $x_{ij}(A) = a_{ij}$.

More generally, if N is an open subset of a manifold M, then N becomes a manifold by restricting the topology and charts of M to N, and N is called an **open submanifold** of M.

(2) The unit circle $M = S^1 = \{a \in R^2 : \|a\| = 1\}$ with the topology induced from R^2 is a one-dimensional manifold, and the collection of open sets which covers S^1 can be taken to have two elements. More generally, we shall show later that the n-sphere $S^n = \{a \in R^{n+1} : \|a\| = 1\}$ is an n-dimensional manifold, and the collection of charts can be taken to have two elements.

(3) The closed interval $M = [0, 1]$ is not a one-dimensional manifold since the point 0 is not contained in an open set $U \subset M$ which is homeomorphic to an open set in R. Is the loop M as indicated in Fig. 2.1 a manifold? Thus is the point of intersection contained in an open set $U \subset M$ which is homeomorphic to an open set in R?

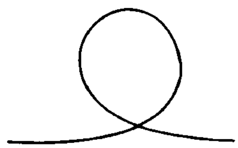


Fig. 2.1.

The coordinate functions given for the manifold $GL(V)$ are differentiable of class C^∞ (actually analytic), and we now define such notions in general.

Definition 2.3 A set $\mathscr{A}$ of (m-dimensional) charts of an m-dimensional manifold M is called a **C^∞-atlas** if $\mathscr{A}$ satisfies the following conditions.

(a) For every $p \in M$, there exists a chart $(U, x) \in \mathscr{A}$ with $p \in U$; that is, $M = \bigcup\{U : (U, x) \in \mathscr{A}\}$.

(b) If $(U(x), x)$ and $(U(y), y)$ are in $\mathscr{A}$, where $U(z)$ is the coordinate neighborhood corresponding to the homeomorphism z, then $U(x) \cap U(y)$ is empty or the maps $x \circ y^{-1}$ and $y \circ x^{-1}$ are of class C^∞.

Note that $x \circ y^{-1}$ (respectively $y \circ x^{-1}$) has domain $y(U(x) \cap U(y))$ [respectively $x(U(x) \cap U(y))$] and transforms these subsets of R^m homeomorphically onto each other (Fig. 2.2). Thus since one of these maps is the inverse

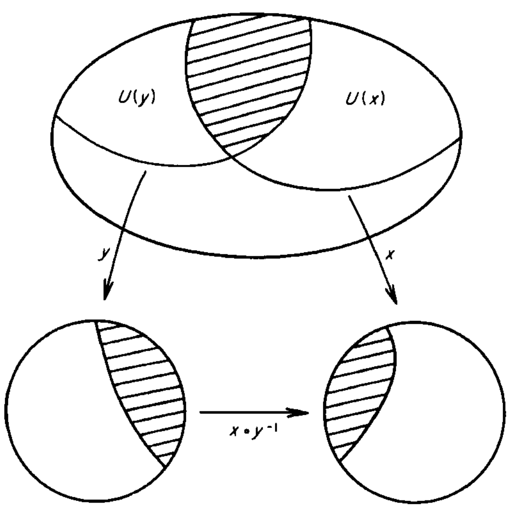


Fig. 2.2.

of the other, their derivatives are invertible linear transformations on R^m, using the chain rule. These maps are called a **change of coordinates** and one says that the corresponding coordinate $(U(x), x)$ and $(U(y), y)$ systems of p are **compatible** when they satisfy condition (b). This will eventually lead to the fact that if a function $f: M \to R$ is differentiable in one coordinate system, then f is differentiable in any compatible coordinate system.

Definition 2.4 Let $\mathscr{A}$ be a C^∞-atlas on an m-dimensional manifold M. Then a chart (U, x) is **admissible to $\mathscr{A}$** or **compatible with $\mathscr{A}$** if (U, x) is compatible with every chart in $\mathscr{A}$; that is, for any $(U(y), y) \in \mathscr{A}$, we have (U, x) and $(U(y), y)$ satisfy condition (b) in Definition 2.3.

Now given any atlas $\mathscr{A}$, one can adjoin all charts which are admissible to $\mathscr{A}$ and obtain a collection $\bar{\mathscr{A}}$ which is again an atlas on M. Thus $\bar{\mathscr{A}}$ is **maximal** relative to properties (a) and (b) of Definition 2.3, and any atlas is contained in a unique maximal atlas.

Definition 2.5 (a) An m-dimensional topological manifold M has a **C^∞-differentiable structure** or just a **C^∞-structure** if one gives M a maximal C^∞-atlas. Thus to give a C^∞-differentiable structure, one need only exhibit a C^∞-atlas on M, then consider the maximal atlas containing it.

(b) A **differentiable manifold of class C^∞** or just a **C^∞-manifold** is an m-dimensional topological manifold M to which there is assigned a C^∞-differentiable structure.

REMARKS (1) One obtains differentiable manifolds of class C^k, $k \geq 0$, or real analytic manifolds by just demanding that the change of coordinates $y \circ x^{-1}$ and $x \circ y^{-1}$ given in Definition 2.3(b) is of class C^k or analytic.

(2) To define an **m-dimensional complex manifold** just replace R^m in the definition of differentiable manifold of class C^∞ by the m-dimensional complex space C^m. Condition (b) in Definition 2.3 must be modified by demanding that the functions $y \circ x^{-1}$ and $x \circ y^{-1}$ be holomorphic in the respective sets in C^m.

Examples (4) Let $M = R$ and define a coordinate system $(U(x), x)$ by $U(x) = R$ and $x : M \to R : t \to t$. Then $\mathscr{A} = \{(U(x), x)\}$ is a C^∞-atlas which defines a differentiable structure and R is a differentiable manifold of class C^∞ relative to this structure. Now let $M_1 = R$ and define a coordinate system $(U(y), y)$ by $U(y) = R$ and $y : M_1 \to R : t \to t^3$. Then $\mathscr{A}_1 = \{(U(y), y)\}$ is a C^∞-atlas since $U(y)$ covers M_1 and the map $y \circ y^{-1}$, the identity, is of class C^∞. Thus Definition 2.3 is satisfied. The atlas $\mathscr{A}_1$ makes M_1 into a C^∞-manifold. The manifolds are distinct in the sense that the charts $(U(x), x)$ and $(U(y), y)$

on R are not compatible since $x \circ y^{-1} : R \to R : t \to t^{1/3}$ is not differentiable at $t = 0$.

(5) Let $S^n = \{a \in R^{n+1} : \|a\| = 1\}$ be the n-sphere with the topology induced from R^{n+1} and $\|a\|^2 = \sum_{i=1}^{n+1} a_i^2$ for $a = (a_1, \ldots, a_{n+1}) \in R^{n+1}$. We define a differentiable structure on S^n as follows. Let $p = (0, \ldots, 0, 1)$ be the "north pole" and $q = (0, \ldots, 0, -1)$ be the "south pole" of S^n. Then the open sets $U(p) = S^n - \{p\}$ and $U(q) = S^n - \{q\}$ cover S^n, and we define coordinate functions x and y so that $\{(U(p), x), (U(q), y)\}$ is an atlas on S^n. The functions x and y are defined by stereographic projections as follows. For $a \in U(p)$ let λ be the line determined by the points p and a and let π be the plane in R^{n+1} given by $u_{n+1} = 0$. Then the value $x(a)$ is the point in R^{n+1} where λ and π intersect. Thus we have a map $x : U(p) \to R^n$ (see Fig. 2.3).

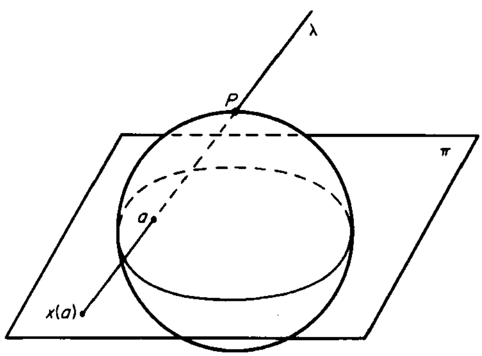


Fig. 2.3.

More specifically if $a = (a_1, \ldots, a_{n+1})$, then $x(a) = (x_1, \ldots, x_n)$, where $x_i = a_i/1 - a_{n+1}$ for $i = 1, \ldots, n$. Similarly y is given by stereographic projection $y : U(q) \to R^n : a \to (y_1, \ldots, y_n)$, where $y_i = a_i/1 + a_{n+1}$ for $i = 1, \ldots, n$. From the formulas, the functions x and y are homeomorphisms onto R^n, and the formulas show that $x \circ y^{-1}$ and $y \circ x^{-1}$ are of class C^∞. Thus we obtain an atlas which makes S^n into a C^∞-manifold.

Note that S^n is a special case of manifolds defined by the implicit function theorem as follows. Let $f: R^{n+1} \to R$ be a C^∞-function and suppose that on the set $M = \{p \in R^{n+1} : f(p) = 0\}$ we have $Df(p) \neq 0$ or more generally $D_{n+1}f(p) \neq 0$. Then one can apply the implicit function theorem to obtain a neighborhood of $p \in M$ which projects in a bijective manner onto the plane $u_{n+1} = 0$ and yields an atlas which makes M into a C^∞ n-dimensional manifold. Thus for S^n, take $f(x_1, \ldots, x_{n+1}) = x_1^2 + \cdots + x_{n+1}^2 - 1$ and note $Df(p) \neq 0$ for $p \in S^n = f^{-1}(0)$ (just compute $Df(p)$ for $p \in S^n$).

(6) We now consider the product manifold determined by two C^∞-manifolds M and N. Thus let $(U(x), x)$ and $(V(y), y)$ be in the maximal atlases

for M and N with $U(x)$ [respectively $V(y)$] a neighborhood of $p \in M$ (respectively $q \in N$). Then define an atlas on the topological product space $M \times N$ by letting $U(x) \times V(y)$ be the coordinate neighborhood of $(p, q) \in M \times N$ and define the homeomorphism

$$x \times y : U(x) \times V(y) \to R^m \times R^n : (u, v) \to \big(x(u), y(v)\big).$$

Thus the set of all these charts $(U(x) \times V(y), x \times y)$ defines a C^∞-atlas on $M \times N$ and the corresponding maximal atlas defines a C^∞-differentiable structure on $M \times N$. The **product manifold of** M **and** N is the Hausdorff space $M \times N$ with the C^∞-structure as given above. Similarly one can define the product of any finite number of differentiable manifolds.

Next let S^1 be the unit circle with the usual C^∞-differentiable structure and let $T^n = S^1 \times \cdots \times S^1$ (n-times) be the product manifold. Then T^n is called an ***n*-dimensional torus.** Thus, in particular, since $T^2 = \bigcup \{\{x\} \times S^1 : x \in S^1\}$; that is, T^2 is a union of unit circles whose centers are on a unit circle, we obtain Fig. 2.4.

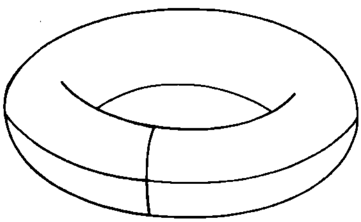


Fig. 2.4.

As shown in Fig. 2.5, T^2 can also be represented as a closed square whose points on the top edge are identified with those directly below on the bottom edge. The points on the right and left edges with the same heights are identified; in particular, the four vertices are identified as the same point. This identification comes from appropriately cutting and bending the above diagram for T^2. Note that since $S^1 = \{\exp 2\pi ix : 0 \le x \le 1\}$, we can identify $T^2 = \{(\exp 2\pi ix, \exp 2\pi iy) : 0 \le x < 1 \text{ and } 0 \le y < 1\}$ with $[0, 1) \times [0, 1)$ as above.

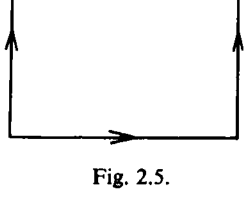


Fig. 2.5.

2. Differentiable Functions

A mapping $f: M \to N$ of two C^∞-manifolds will be seen to be differentiable of class C^∞ if its "coordinate expressions" are differentiable. Thus we shall reduce the differentiability of f to investigating the differentiability of functions $g: R^m \to R$.

Definition 2.6 Let M and N be C^∞-manifolds of dimension m and n, respectively, and let

$$f: M \to N$$

be a map defined on a neighborhood of a point $p \in M$. We say that f is **differentiable at** p **of class** C^∞ if there exists a coordinate system (U, x) at p in M and a coordinate system (V, y) at $f(p)$ in N such that

$$y \circ f \circ x^{-1} : x(U) \to y(V)$$

is differentiable at $x(p)$ of class C^∞ (see Fig. 2.6). Note that $x(U) \subset R^m$ and $y(V) \subset R^n$.

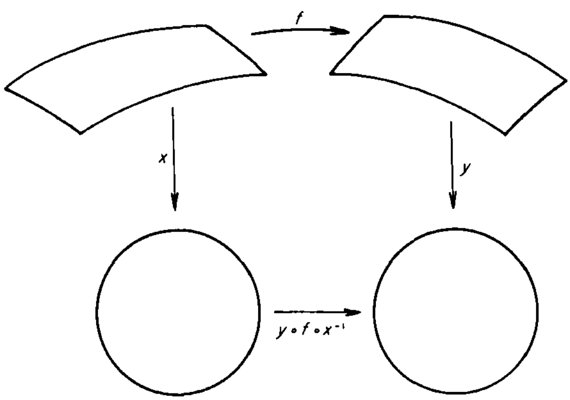


Fig. 2.6.

Since differentiability is given in terms of specific charts, we must show that it is actually independent of the choice of charts. Thus let $(\bar{U}, \bar{x})$ [respectively $(\bar{V}, \bar{y})$] be any other elements of the atlas for M (respectively N) which are neighborhoods of p [respectively $f(p)$]. Then we must show the map

$$\bar{y} \circ f \circ \bar{x}^{-1} : \bar{x}(\bar{U}) \to \bar{y}(\bar{V})$$

is differentiable at $\bar{x}(p)$. However, since differentiability is a local property, it suffices to show this on neighborhoods. Thus for $U \cap \bar{U}$ and $V \cap \bar{V}$ (which are nonempty), we have on the neighborhoods $\bar{x}(U \cap \bar{U})$ and $\bar{y}(V \cap \bar{V})$ that

$$\bar{y} \circ f \circ \bar{x}^{-1} = \bar{y} \circ y^{-1} \circ (y \circ f \circ x^{-1}) \circ x \circ \bar{x}^{-1}.$$

Thus since $\bar{y} \circ y^{-1}$, $y \circ f \circ x^{-1}$ and $x \circ \bar{x}^{-1}$ are of class C^∞, so is their composition. If $f: M \to N$ is C^∞-differentiable at every point $p \in M$, then f is a **differentiable map of class C^∞ from M into N.**

Now in terms of coordinates, it suffices to show that the functions

$$f_i = u_i \circ (y \circ f \circ x^{-1}) : R^m \to R$$

for $i = 1, \ldots, n$ are differentiable on an open subset D of $x(p)$. Thus if (U, x) and (V, y) are the corresponding coordinate systems, then we obtain the coordinate expression

$$y_i = f_i(x_1, \ldots, x_m) \qquad \text{for} \quad i = 1, \ldots, n$$

which must be differentiable at $x(p) = (p_1, \ldots, p_m)$. This yields the following; for example, see Bishop and Goldberg [1968, p. 37].

Proposition 2.7 Let $f: M \to N$ be a continuous mapping of two C^∞-manifolds. Then f is of class C^∞ on M if and only if for every real-valued C^∞-function $y: V \to R$ defined on an open submanifold V of N, the function $y \circ f$ is of class C^∞ on the open submanifold $f^{-1}(V)$ of M.

We shall write $C^\infty(M)$ or $F(M)$ for the set of real-valued C^∞-functions on M and $C^\infty(p)$ or $F(p)$ for the set of those real-valued functions which are C^∞-differentiable at $p \in M$. Note that since differentiability of f at $p \in M$ also means f is defined on a neighborhood U of p, the elements of $C^\infty(p)$ are actually pairs (f, U). Consequently one can define an equivalence relation for elements of $C^\infty(p)$ such that $(f_1, U_1) \sim (f_2, U_2)$ if and only if there exists an open set G with $p \in G$ and $f_1(q) = f_2(q)$ for all $q \in G$. The set of equivalence classes are called **germs of C^∞-differentiable functions at p.** Note that the coordinate functions x_i on U are in $C^\infty(p)$. We shall usually not use this terminology but just the underlying ideas.

Next note that $F = C^\infty(M)$ is an associative algebra over R with operations given by

$$\begin{aligned} (af)(p) &= af(p) && \text{for} \quad a \in R, \\ (f + g)(p) &= f(p) + g(p) && \\ (fg)(p) &= f(p)g(p) && \text{for} \quad f, g \in C^\infty(M) \end{aligned}$$

and F satisfies the following [Helgason, 1962, p. 5].

(1) If $f_1, \ldots, f_r \in F$ and if $g: R^r \to R$ is of class C^∞ on R^r, then $g(f_1, \ldots, f_r) \in F$.

(2) If $f: M \to R$ is a function on M such that for each $p \in M$ there is a $g \in F$ and there is a neighborhood U of p such that $f(q) = g(q)$ for all $q \in U$, then $f \in F$.

(3) For each p in the m-dimensional manifold M, there exist m functions $f_1, \ldots, f_m$ in F and an open neighborhood U of p such that the mapping

$$U \to R^m : q \to (f_1(q), \ldots, f_m(q))$$

is a homeomorphism of U onto an open subset of R^m. The functions $f_1, \ldots, f_m$ and the set U can be chosen so that for any $f \in F$, there is $g : R^m \to R$ of class C^∞ and

$$f = g(f_1, \ldots, f_m)$$

on U.

These properties determine a differentiable structure on M as follows (see Helgason [1962, p. 6] for a proof).

Proposition 2.8 Let M be a topological Hausdorff space and let m be an integer greater than 0. Let F be a set of real-valued functions on M satisfying properties (1)–(3). Then there exists a unique collection of charts $\mathscr{A} = \{(U_\alpha, x_\alpha) : \alpha \in A\}$ which form a maximal atlas of M such that the set of real-valued C^∞-functions on the manifold M with atlas $\mathscr{A}$ equal the set F.

Definition 2.9 The C^∞-manifolds M and N are **diffeomorphic** if there exists a homeomorphism $f: M \to N$ such that f and f^{-1} are of class C^∞; f is called a **diffeomorphism.**

Thus a diffeomorphism yields an equivalence relation such that the two manifolds are not only topologically equivalent, but also they have equivalent differentiable structures.

Examples (1) Let R be a manifold with the usual structure $x : R \to R : t \to t$ and $(-1, 1)$ be an open submanifold of R. Then

$$f: (-1, 1) \to R : t \to t/(1 - t^2)$$

is a diffeomorphism.

(2) Let R be the above manifold with the usual structure, and let M_1 be the manifold with space R and coordinate function $y: M_1 \to R : t \to t^3$. Then the map $f: M_1 \to R : s \to s^3$ is a C^∞-homeomorphism and the inverse homeomorphism $f^{-1}: R \to M_1 : u \to u^{1/3}$ is actually differentiable of class C^∞ relative to the above differentiable structures: For $t \in R$ we have the coordinate expression $(y \circ f^{-1} \circ x^{-1})(t) = (y \circ f^{-1})(t) = y(t^{1/3}) = t$. However, note

that identity map $g : M_1 \to R : t \to t$ is not C^∞ since $(x \circ g \circ y^{-1})(t) = t^{1/3}$ which is not C^∞, that is, the identity map is not a diffeomorphism.

Exercise Let M be a C^∞-manifold. Show that the charts (U, x) and (V, y) at $p \in M$ are compatible if and only if x and y are C^∞-related by $y = f(x)$ and $x = g(y)$ for suitable C^∞-functions f and g.

3. Submanifolds

We shall now use the preceding results to study certain substructures of a manifold and return to these topics again after studying the differential of a function.

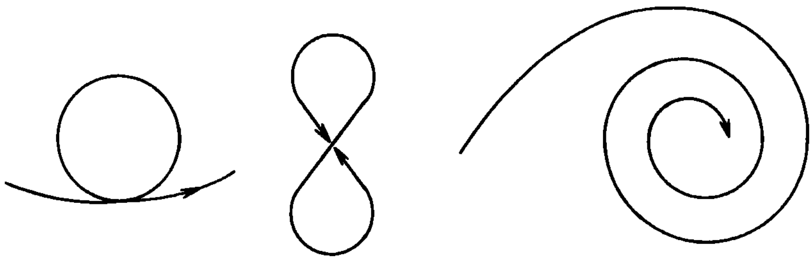
Definition 2.10 Let M and N be C^∞-manifolds of dimensions m and n respectively, and let $f: M \to N$ be a C^∞-mapping.

(a) We call f an **immersion** of M into N if for every $p \in M$, there is a neighborhood U of p in M and a chart (V, y) of $f(p)$ in N such that if we write $y = (y_1, \ldots, y_n)$ in terms of coordinate functions, then $x_i = y_i \circ f \mid U$ for $i = 1, \ldots, m$ are coordinate functions on U in M. That is if $x = (x_1, \ldots, x_m)$, then (U, x) is a chart at p in M. We say that M is **immersed** in N if an immersion $f: M \to N$ exists.

(b) We call f an **embedding** if f is injective and f is an immersion. Also M is said to be **embedded** in N. Thus an immersion is a local embedding.

(c) We call the subset $f(M)$ of N a **submanifold** of N if f is an embedding and $f(M)$ is given a C^∞-differential structure for which the mapping of manifolds $f: M \to f(M)$ is a diffeomorphism. In particular if M is a subset of the C^∞-manifold N and M has its own C^∞-differentiable structure, then M is a submanifold of N if the inclusion map $i : M \to N : x \to x$ is an embedding. Thus a coordinate system on N induces a coordinate system on M.

The subset $f(M)$ is called an **immersed submanifold** if the above mapping f is just an immersion. The topology of a submanifold $M \subset N$ need not be the induced topology of the containing manifold. However, since the inclusion map is C^∞ and consequently continuous, the open sets in the induced topology are open sets in the submanifold topology. Also note that the dimension of a submanifold is less than or equal to the dimension of its containing manifold and in the case of equality we just obtain open submanifolds; this can be easily seen by using the inverse function theorem as stated in Section 2.5.



(a) (b) (c)

Fig. 2.7.

Examples (1) Consider the mappings $f\colon R \to R^2$ indicated in Fig. 2.7. In (a) f is an immersion (why?) but not an embedding and $f(R)$ is an immersed submanifold but not a submanifold. In (b) the figure "8" is such that the arrow segments approach but do not touch the center p. Then f is an embedding and $f(R)$ is a submanifold when given the obvious C^∞-structure. Note that the submanifold topology is that of a bent open interval and therefore a neighborhood of p in the submanifold topology is just a bent open interval containing p. However, a neighborhood of p in the topology induced from R^2 always contains part of the arrow curves near p. Also the spiral in (c) yields an embedding and a submanifold. What can be said about the submanifold topology and the topology induced from R^2 in (c)?

(2) Consider the torus of Section 2.1

$$T^2 = \{(\exp 2\pi ix, \exp 2\pi iy) : 0 \leq x < 1 \text{ and } 0 \leq y < 1\}$$

and define

$$f\colon R \to T^2 : t \to (\exp 2\pi iat, \exp 2\pi ibt)$$

where $a/b = \alpha$ is an irrational number. Then f is injective (by solving the resulting equations and using α is irrational) and f is C^∞. Thus by giving $f(R)$ the obvious C^∞-structure so that $f\colon R \to f(R)$ is a diffeomorphism, $f(R)$ is a submanifold. Furthermore $f(R)$ wraps around T^2 in a nonintersecting manner and is actually dense in T^2 (exercise or see the text of Auslander and MacKenzie [1963]). Representing T^2 as a square with opposite sides identified as discussed in Section 2.1, we see $f(R)$ can be represented by the line segment $(x, y) = (at, bt)$ and their displacements as in Fig. 2.8. Also we should note

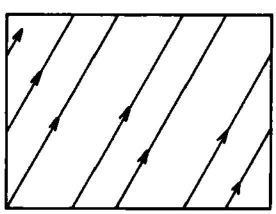


Fig. 2.8.

that points close together in T^2 need not be close in $f(R)$; that is, the topology in $f(R)$ is not the induced topology.

Exercises (1) In general, can one find a one-dimensional submanifold of T^n which is dense in T^n?

(2) Show if $f: M \to N$ defines a submanifold and if M is compact, then $f: M \to f(M)$ is a homeomorphism. (*Hint*: What can be said about a continuous map of a compact space onto a Hausdorff space?)

If $z = f(x, y)$ is a well-behaved function, then it defines a surface $M \subset R^3$ which is a two-dimensional submanifold. For a point $p \in M$ we can define, in a suitable neighborhood V in R^3 of p, coordinates (x, y, u), where $u = z - f(x, y)$. Thus the surface is given locally by the equation $u = 0$. The familiar upper hemisphere given by $z = (1 - x^2 - y^2)^{1/2} > 0$ is an example of such a situation. We have the following generalization of this.

Proposition 2.11 Let M be an m-dimensional C^∞-submanifold of the n-dimensional C^∞-manifold N and let $p \in M$. Then there exists a coordinate system (V, z) of N with $p \in V$ such that:

(a) $z_1(p) = \cdots = z_n(p) = 0$ where the z_i are the coordinate functions;

(b) the set $W = \{r \in V: z_{m+1}(r) = \cdots = z_n(r) = 0\}$ together with the restriction of $z_1, \ldots, z_m$ to W form a chart of M with $p \in W$.

Conversely, if a subset $M \subset N$ has a manifold structure with a coordinate system at each $p \in M$ satisfying the above, then M is a submanifold of N.

PROOF Let $f: Q \to N$ be an embedding which defines $M = f(Q)$ and let $p = f(q)$ for a unique $q \in Q$. Now let (T, y) be a chart for p in N and we can assume $y(p) = 0$ in R^n. Let U be a neighborhood of $q = f^{-1}(p)$ in Q and let $x = y \circ f | U$ be such that (U, x) is a chart for q in Q. Thus $x(q) \in R^m$ and for $i = 1, \ldots, m$ we have $x_i = y_i \circ f | U$ are the corresponding coordinate functions.

Now the composition $y \circ f \circ x^{-1} = F$ defines a C^∞-function $F: x(U) \to y(T)$, where $x(U) \subset R^m$ and $y(T) \subset R^n$, and we can write F in terms of coordinates

$$y_i = f_i(x_1, \ldots, x_m) \qquad \text{for} \quad i = 1, \ldots, n.$$

The hypotheses $M = f(Q)$ is a submanifold, $y \circ f = F \circ x$, and $x = y \circ f | U$ yield $y_i = x_i$ for $i = 1, \ldots, m$ in the above expression for F. Thus the rank of $DF(x(q))$ is m; that is, the $m \times m$ matrix $(\partial f_i/\partial x_j)$, $i, j = 1, \ldots, m$, is the identity. By the inverse function theorem there exists a neighborhood D of $x(q)$ with $D \subset x(U)$ where the first m equations can be locally inverted

$$x_i = g_i(y_1, \ldots, y_m) \qquad \text{for} \quad i = 1, \ldots, m,$$

where $y_1, \ldots, y_m$ are actually the coordinate functions defined on $f \circ x^{-1}(D) \subset T$ but are also used above to denote "coordinates" in $y(T)$. (Note that due to the simple expression of the first m equations $y_i = f_i(x_1, \ldots, x_m)$, the functions g_i can be explicitly computed. What are they?)

We now change from the y-coordinates to the $z = (z_1, \ldots, z_n)$ coordinates given by

$$z_j = y_j \qquad \text{for} \quad j = 1, \ldots, m$$
$$z_i = y_i - f_i(g_1(y_1, \ldots, y_m), \ldots, g_m(y_1, \ldots, y_m))$$

for $i = m + 1, \ldots, n$. These equations for y's are defined on $f \circ x^{-1}(D) \subset T$ and are C^∞. They form a change of coordinates because z^{-1} exists locally [show $\det(\partial z_i/\partial y_j(p)) \neq 0$] and $y \circ z^{-1}$ and $z \circ y^{-1}$ are C^∞ on their domains.

Let $V \supset f \circ x^{-1}(D)$ be the subset of the domain of y in T where z is defined. Then V is a neighborhood of p in N and by unscrambling the definitions of f_i and g_i and using $y(p) = 0$ we have $z(p) = 0$. Now the set W in (b) given by

$$W = \{r \in V : z_{m+1}(r) = \cdots = z_n(r) = 0\}$$

contains p and is in M since in terms of the defining equations for $z_{m+1}, \ldots, z_n$ we see $W \subset f(x^{-1}(D)) \subset f(U) \subset f(Q) = M$ and also W is open in M. The restriction of $z_i = y_i$ for $i = 1, \ldots, m$ to W equals x_i (second paragraph) and so are coordinates on W.

The converse follows from various definitions.

Remark The above proof contains some machinery which is not necessary in view of our definition of a submanifold and for a more direct proof see the book by Bishop and Goldberg [1968, p. 42]. However, it can be modified to obtain the following result which is frequently used as the definition of a submanifold [Helgason, 1962; Singer and Thorpe, 1967].

Corollary 2.12 Let P be an m-dimensional C^∞-manifold, let N be an n-dimensional C^∞-manifold with $n \geq m$, and let $f: P \to N$ be an injective C^∞-function. If for every $q \in P$, there exists a chart (U, x) of q in P and there exists a chart (T, y) of $f(q) = p$ in N such that the linear transformation

$$D(y \circ f \circ x^{-1})(x(q)) : R^m \to R^n$$

is injective, then $M = f(P)$ is a submanifold of N provided $f(P)$ is given a C^∞-structure so that $f: P \to f(P)$ is a diffeomorphism.

Proof We shall use the converse of Proposition 2.11 by showing (a) and (b) hold. By a simple translation argument we can assume that $x(q) = y(p) = 0$. Now near $x(q)$ we can represent the composition $y \circ f \circ x^{-1} = F$ in terms of coordinates

$$y_i = f_i(x_1, \ldots, x_m) \qquad \text{for} \quad i = 1, \ldots, n.$$

Since $D(y \circ f \circ x^{-1})(x(q))$ has rank m we have that some subsystem of m equations

$$y_{i_j} = f_{i_j}(x_1, \ldots, x_m) \qquad \text{for} \quad j = 1, \ldots, m$$

is such that $m \times m$ matrix $(\partial f_{i_j}/\partial x_k)$ is invertible. We can assume this subsystem consists of the first m equations and consequently can define a function

$$\bar{F} : x(U) \to R^m : (x_1, \ldots, x_m) \to (y_1, \ldots, y_m)$$

where we use the coordinate function to also denote the corresponding point. Thus by the inverse function theorem there exists a neighborhood D of $x(q)$ and $D \subset x(U)$ on which $\bar{F}$ has a local inverse G. Thus the first m equations can be locally inverted

$$x_i = g_i(y_1, \ldots, y_m) \qquad \text{for} \quad i = 1, \ldots, m,$$

and we proceed as in the above proof. Note that from the defining equations of z we see that $x(q) = y(p) = 0$ implies $z(p) = 0$.

REMARK Let M be an m-dimensional C^∞-manifold. Then it can be proved that M is diffeomorphic to a submanifold of R^n with $n \leq 2m + 1$. This theorem of Whitney can be found in the work of Auslander and MacKenzie [1963].

Proposition 2.13 Let M and N be C^∞-manifolds of dimension m and n, respectively, with $m \geq n$. Let $f: M \to N$ be a C^∞-map and for some fixed $p \in N$ let $f^{-1}(p) = \{q \in M : f(q) = p\}$. Let every $q \in f^{-1}(p)$ have a chart (U, x) in M and let p have a chart (T, y) in N such that $D(y \circ f \circ x^{-1})(x(q)) : R^m \to R^n$ is surjective. Then $f^{-1}(p)$ is a closed $(m - n)$-dimensional submanifold of M or $f^{-1}(p)$ is empty.

PROOF This follows from the variation of the inverse function theorem given in Proposition 1.17 using the inverse image of the set $\{p\}$ is closed (or see the book by Spivak [1965, p. 111]).

A C^∞-map $f: M \to N$ such that for every $q \in M$ there is a chart (U, x) at q and a chart (T, y) at $f(q)$ with $D(y \circ f \circ x^{-1})(x(q))$ surjective is called a **submersion**. Thus the injective or surjective nature of $D(y \circ f \circ x^{-1})(x(q))$ determines submanifolds.

Exercise (3) If $f: R \to R^m$ is C^∞, then show the graph of f given by $G(f) = \{(t, f(t)) : t \in R\}$ is a submanifold of $R^{m+1} = R^1 \times R^m$ with the induced topology. Does $f: R \to R^2 : t \to (t^2, t^3)$ define a submanifold? The above can be generalized to C^∞-functions $f: M \to N$.

Definition 2.14 Let M be a C^∞-manifold. A C^∞**-curve** in M is a C^∞-map f from some interval I contained in R into M such that f has an extension $\tilde{f}$ which is a C^∞-map of an open interval $J \supset I$ into M. Thus if $I = [a, b]$, then there exists an $\varepsilon > 0$ such that $J = (a - \varepsilon, b + \varepsilon)$ and there exists a C^∞-function $\tilde{f}: J \to M$ such that $f(t) = \tilde{f}(t)$ for all $t \in I$. Then f is frequently called a **curve segment** in case $I = [a, b]$. A **broken** C^∞**-curve** in M is a continuous map f: $[a, b] \to M$ together with a partition of $[a, b]$ such that on the corresponding closed subintervals f is a C^∞-curve.

Examples (3) The map $f: R \to R^2 : t \to (t^2, t^3)$ is a C^∞-curve in R^2 with a cusp at (0, 0).

(4) The "wrap around" map on the torus T^2 given in Section 2.3 with "irrational slope" is actually a C^∞-curve which is dense in T^2.

(5) The map f: $[0, 1] \to R^2$ given by

$$f(t) = \begin{cases} (t, \sin 1/t) & \text{if} \quad t \neq 0, \\ (0, 0) & \text{if} \quad t = 0 \end{cases}$$

is not a C^∞-curve in R^2 since it does not have a C^∞-extension to an open interval containing 0.

We recall that a topological space M is **connected** if it satisfies any of the following equivalent conditions:

(1) M is not the union of two nonempty disjoint closed subsets;

(2) M is not the union of two nonempty disjoint open subsets;

(3) the only subsets of M which are both open and closed are M and the empty set;

(4) if $M = \bigcup_a E_a$, where E_a are open and $E_a \cap E_b$ is empty if $a \neq b$, then only one of the E_a is nonempty;

(5) if $f: M \to N$ is a continuous map into a discrete set, then $f(M)$ is a single point.

A topological space M is **path connected** if for every $p, q \in M$, there exists a continuous curve f: $[a, b] \to M$ with $p = f(a)$ and $q = f(b)$. We have the fact that a path connected space must be connected [Singer and Thorpe, 1967].

Proposition 2.15 Let M be a C^∞-manifold.

(a) If M is connected, then every pair of points can be joined by a broken C^∞-curve.

(b) M is connected if and only if M is path connected.

Proof Part (b) follows from the preceding remarks and part (a). Thus let $p \in M$ and for $q \in M$, define $q \sim p$ if and only if q can be joined to p by a broken C^∞-curve. Then since $\sim$ is an equivalence relation, M is the union of the disjoint equivalence classes

$$E_p = \{q \in M : q \sim p\}.$$

Now each E_p is open in M, for if $q \in E_p$, let (U, x) be a chart of M such that $q \in U$ with $x(q) = 0$ and $x(U) = B_m$, an open m-ball. Now for each $u \in U$ the point $x(u) \in B_m$ can be joined to $x(q)$ by a C^∞-curve λ in B_m; that is, λ a straight line segment. Therefore $u \in U$ can be joined to q by the C^∞-curve $x^{-1} \circ \lambda$ and consequently $u \in U$ can be joined to p by a broken C^∞-curve. Thus $u \sim p$, so that $U \subset E_p$ and E_p is open. However, since $M = \bigcup E_q$ (disjoint), we have by condition (4) that all the E_q are empty except one. Thus every point in M can be joined to p by a broken C^∞-curve.

Example (6) Let $V = R^n$ and let $G = GL(V)$. Then G is an open n^2-dimensional submanifold of R^{n^2}. Now let

$$SL(V) = \{A \in GL(V) : \det(A) = 1\}.$$

Then $SL(V)$ is clearly a subgroup of $GL(V)$ and is called the **special linear group** and is sometimes denoted by $SL(n, R)$. Now $SL(V)$ is a closed submanifold because if we let

$$f : GL(V) \to R - \{0\} : A \to \det(A),$$

then using exercise (5), Section 1.4 for the derivative of det we see that for all $A \in GL(V)$, $D(f)(A)$ is surjective; that is, of rank 1. Thus by Proposition 2.13, $SL(V) = f^{-1}(1)$ is closed and of dimension $n^2 - 1$.

We shall now use exp to obtain a coordinate system at the point $I \in SL(V)$ and then for any point $A \in SL(V)$. For $g = gl(V)$ we let

$$sl(V) = \{X \in g : \operatorname{tr} X = 0\}.$$

Thus since tr is linear and $\operatorname{tr}[X, Y] = \operatorname{tr} XY - \operatorname{tr} YX = 0$ we see that $sl(V)$ is a Lie subalgebra of g; that is, $sl(V)$ is a vector subspace of g so that for all $X, Y \in sl(V)$ we have $[X, Y] = XY - YX \in sl(V)$. Also for any $X \in g$,

$$X = \frac{1}{n}(\operatorname{tr} X)I + \left[X - \frac{1}{n}(\operatorname{tr} X)I\right] = \frac{1}{n}(\operatorname{tr} X)I + Y,$$

where $\operatorname{tr} Y = 0$. Consequently $\dim sl(V) = n^2 - 1$. Next note that exp restricted to $sl(V)$ is actually in $SL(V)$, since we have from exercise (3), Section 1.1 that

$$\det(\exp X) = e^{\operatorname{tr} X} = 1$$

if $X \in sl(V)$. Thus if we let $F = \exp | sl(V)$, we have from the proof of Proposition 1.19 that $DF(0)$ is the identity. Therefore by the inverse function theorem there exists a neighborhood U_0 of 0 in $sl(V)$ and a neighborhood U_I

of I in $SL(V)$ such that $F: U_0 \to U_I : X \to \exp X$ is a diffeomorphism of U_0 onto U_I. Thus for t in a sufficiently small interval $(-\delta, \delta)$ of $0 \in R$ and for X fixed in $sl(V)$ we see that the map

$$\exp : sl(V) \to SL(V) : tX \to \exp tX$$

maps the line segment tX into a C^∞-curve segment in $SL(V)$.

To coordinatize $SL(V)$ by exp we proceed as follows. First as in the remarks following Proposition 1.19 we have the local C^∞-diffeomorphism

$$\log : U_I \to U_0$$

and since U_0 is open in $sl(V)$ we find that $(U_I, \log)$ is a chart at I in $SL(V)$ [noting that $\operatorname{tr}(\log \exp X) = 0$]. Now for any other point $A \in SL(V)$ we have that

$$L(A) : SL(V) \to SL(V) : B \to AB$$

is a diffeomorphism of $SL(V)$ and therefore the set

$$L(A)U_I = AU_I = \{Au : u \in U_I\}$$

is an open neighborhood of A, using $A = AI$. Let $V = AU_I$ and let $y = \log \circ L(A)^{-1}$. Then (V, y) is a chart at A in $SL(V)$ as shown in Fig. 2.9. Finally we remark that $G = GL(V)$ is not connected; for if it were, then since det is continuous, $\det(G) = R - \{0\}$ is connected, a contradiction. However, $SL(V)$ is connected and this follows from Proposition 2.15 and the following result.

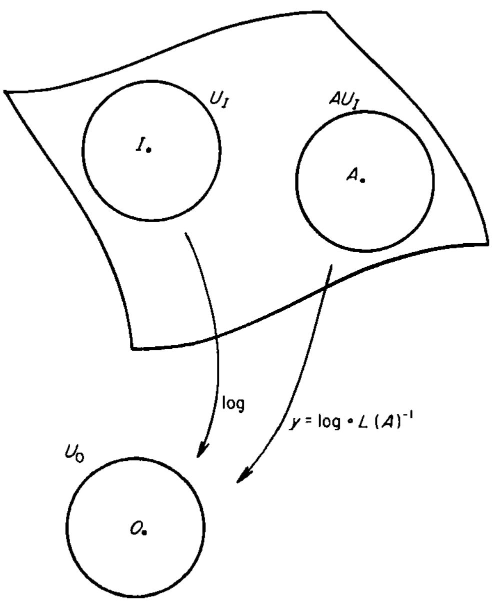


Fig. 2.9.

Exercise (4) Let $P(V) = \{A \in GL(V) : \det A > 0\}$. Then $P(V)$ is path connected.

Now to show $SL(V)$ is connected we note that the map

$$\theta : P(V) \to SL(V) : A \to (\det A)^{-1/n} A$$

is a continuous surjection so that $SL(V)$ is connected.

Example (7) Again let $V = R^n$ and let

$$B : V \times V \to R : (X, Y) \to B(X, Y)$$

be a nondegenerate bilinear form (symmetric or skew-symmetric). Then the adjoint A^* relative to B is uniquely given by

$$B(AX, Y) = B(X, A^* Y)$$

for $A \in \mathrm{End}(V)$. We have the usual rules

$$(aA + bB)^* + aA^* + bB^* \qquad \text{and} \qquad (AB)^* = B^*A^*.$$

Let

$$K = \{B \in \mathrm{End}(V) : B^* = B\}.$$

Then K is a vector space and a manifold and the manifold dimension equals the vector space dimension. Also for any $A \in G = GL(V)$ let

$$f : G \to \mathrm{End}(V) : A \to AA^* - I.$$

Then $f(A) \in K$ and let

$$\begin{aligned} H &= \{A \in G : B(AX, AY) = B(X, Y) \text{ all } X, Y \in V\} \\ &= \{A \in G : AA^* - I = 0\} \\ &= f^{-1}(0). \end{aligned}$$

Then H is clearly a subgroup of G and H is a closed submanifold of G of dimension $n^2 - \dim K$ as follows.

To see this we shall use Proposition 2.13. Thus we must show for every $A \in G$ such that $f(A) = 0$ that $Df(A) : \mathrm{End}(V) \to K$ is surjective. For any $A \in G$ and any $Y \in K$, let $X = \frac{1}{2} Y A^{*-1} \in \mathrm{End}(V)$. Then we shall show

$$Y = [Df(A)](X)$$

so that $Df(A)$ is surjective. Thus

$$\begin{aligned} [Df(A)](X) &= \lim_{t \to 0} \frac{1}{t} [f(A + tX) - f(A)] \\ &= \lim_{t \to 0} \frac{1}{t} \{[(A + tX)(A + tX)^* - I] - (AA^* - I)]\} \\ &= XA^* + (XA^*)^* = Y. \end{aligned}$$

To coordinatize H we proceed as follows. Let $g = gl(V)$ and let

$$\begin{aligned} h &= \{P \in g : B(PX, Y) = -B(X, PY) \text{ all } X, Y \in V\} \\ &= \{P \in g : P^* = -P\}. \end{aligned}$$

Then h is a Lie subalgebra of g; that is, h is a vector subspace of g and for $P, Q \in h$ we have $[P, Q] = PQ - QP \in h$. Thus for $P, Q \in h$ and $a, b \in R$ we have

$$(aP + bQ)^* = aP^* + bQ^* = -(aP + bQ)$$

so that h is a subspace and

$$[P, Q]^* = (PQ)^* - (QP)^* = -(PQ - QP)$$

so that $[P, Q] \in h$ as desired.

Now for any $P \in h$ we have for all $X, Y \in V$ that

$$\begin{aligned} B((\exp P)X, (\exp P)Y) &= B(X, (\exp P)^*(\exp P)Y) \\ &= B(X, (\exp P^*)(\exp P)Y) \\ &= B(X, \exp(-P)(\exp P)Y) = B(X, Y). \end{aligned}$$

Thus $\exp : h \to H$ so that as in the preceding example there exist a neighborhood U_0 of 0 in h and a neighborhood U_I of I in H such that $\exp : U_0 \to U_I$ is an analytic diffeomorphism and $(U_I, \log)$ is a chart at I in H which induces the chart $(AU_I, \log \circ L(A)^{-1})$ at A in H. Also if for any $P \in g$ we demand that the C^∞-curve $R \to G : t \to \exp tP$ actually be in H for t in an interval about 0 in R, then by differentiating the formula $B((\exp tP)X, (\exp tP)Y) = B(X, Y)$ we obtain $P \in h$ using example (1), Section 1.2. Note that the manifold dimension of H equals the vector space dimension of h.

There are various subcases depending on B.

***B* Symmetric** (1) (i) Let B be positive definite; that is, $B(X, X) = 0$ implies $X = 0$. Thus there exists a basis $e_1, \ldots, e_n$ of V such that if $X = \sum x_i e_i$, $Y = \sum y_i e_i$, then $B(X, Y) = \sum x_i y_i$. In this case H is called the **orthogonal group** and denoted by $O(n)$. We also note that the vector space $K = \{B \in \text{End}(V) : B = B^*\}$ is just the set of symmetric matrices and has dimension $n(n + 1)/2$. Thus the manifold dimension of $H = O(n)$ is $n^2 - n(n + 1)/2 = n(n - 1)/2$.

Now for $A \in O(n)$, $AA^* = I$ yields $(\det A)^2 = 1$ so that $\det A = \pm 1$. Thus noting

$$A = \begin{bmatrix} -1 & & & \\ & 1 & & 0 \\ & 0 & \ddots & \\ & & & 1 \end{bmatrix} \in O(n),$$

we have, since $\det : O(n) \to R - \{0\}$ is continuous, that $O(n)$ is not connected.

Let

$$SO(n) = \{A \in O(n) : \det A = 1\}$$
$$= O(n) \cap SL(n, R)$$

which is called the **special orthogonal group**. We know that $SO(n)$ is also a manifold of dimension $n(n-1)/2$ and we shall show later that $SO(n)$ is connected. In this case the Lie algebra $h \cap sl(V)$ associated with $SO(n)$ is denoted by $so(n)$.

(ii) Now assume the general form for the nondegenerate form B; that is, there exists a basis $f_1, \ldots, f_n$ of V such that for $X = \sum x_i f_i$, $Y = \sum y_i f_i$, then

$$B(X, Y) = -\sum_{i=1}^{p} x_i y_i + \sum_{i=p+1}^{n} x_i y_i$$

[Jacobson, 1953, Vol. II]. In this case the group $H \cap SL(n, R)$ is frequently denoted by $SO(p, q)$, where $p + q = n$ and the Lie algebra $h \cap sl(n)$ is denoted by $so(p, q)$.

Next we shall consider $V = R^n$ as column vectors with

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

relative to the basis $f_1, \ldots, f_n$ so that we can write $B(X, Y)$ in block form

$$B(X, Y) = X^t \begin{bmatrix} -I_p & 0 \\ 0 & I_q \end{bmatrix} Y \equiv X^t B Y,$$

where t denotes transpose and I_p, I_q are the appropriate identity matrices. Then for

$$P \in so(p, q) = \{P \in sl(n) : B(PX, Y) = -B(X, PY)\}$$

we have

$$0 = (PX)^t B Y + X^t B(PY) = X^t(P^t B + BP) Y.$$

Thus $P^t B + BP = 0$ and if we partition P into appropriate blocks

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

then $P_{11}^t = -P_{11}$, $P_{22}^t = -P_{22}$, $P_{12} = P_{21}^t$, and P_{12} arbitrary. Thus we obtain the form of the Lie algebra $so(p, q)$ and that it is of dimension

$$p(p-1)/2 + q(q-1)/2 + pq = n(n-1)/2$$

using $p + q = n$.

As before we can use $(U_I, \log)$ to coordinatize $SO(p,q)$ where log: $U_I \to U_0 \subset so(p,q)$. Thus we see that $SO(p,q)$ is an $n(n-1)/2$-dimensional manifold.

Exercise (5) Show $SO(p,q)$ is not connected (*Hint*: Investigate matrices of the form

$$\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

where A_{ii} are orthogonal matrices of the appropriate size satisfying $\det A_{11} \det A_{22} = 1$).

B Skew-symmetric (2) Thus $B(X, Y) = -B(Y, X)$ and there exists a basis $f_1, \ldots, f_n$ of V such that $n = 2p$ and using the preceding notation $B(X, Y)$ has the block form [Jacobson, 1953, Vol. II]

$$B(X, Y) = X^t \begin{bmatrix} 0 & I_p \\ -I_p & 0 \end{bmatrix} Y = \sum_{k=1}^{p} (x_k y_{k+p} - x_{k+p} y_k).$$

In this case we shall consider $H \cap GL(2p, R)$ which is freqently denoted by $Sp(p, R)$, where $n = 2p$, or $Sp(p)$, or $Sp(n, R)$ and is called the **symplectic group**. The Lie algebra associated with $Sp(p, R)$ equals $h \cap gl(2p)$ and is denoted by $sp(p, R)$. Next for $P \in sp(p, R)$ we put it into block form and find

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

where $P_{22} = -P_{11}^t$, $P_{12}^t = P_{12}$, $P_{21}^t = P_{21}$, and P_{11} arbitrary. Thus $p^2 + p(p+1)/2 + p(p+1)/2 = 2p^2 + p = \dim sp(p, R)$ which equals the manifold dimension of $Sp(p, R)$.

For future reference we present a short list of important Lie groups and Lie algebras in Tables 2.1 and 2.2. We will describe the groups and algebras entirely in terms of matrices. For convenience we include the groups and algebras that have been previously discussed. First define $I_{p,q} \in GL(p+q, R)$ and $J_n \in GL(2n, R)$ by

$$I_{p,q} = \begin{bmatrix} -I_p & 0 \\ 0 & I_q \end{bmatrix}, \qquad J_n = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

In the definition of the **unitary group** $U(n)$ the matrix $\bar{X} = (\bar{a}_{ij})$ is the complex conjugate matrix of $X = (a_{ij})$.

TABLE 2.1

LIE GROUPS

$GL(n, C)$	nonsingular $n \times n$ complex matrices,
$GL(n, R)$	nonsingular $n \times n$ real matrices,
$SL(n, C)$	$\{X \in GL(n, C) : \det(X) = 1\}$,
$SL(n, R)$	$\{X \in GL(n, R) : \det(X) = 1\}$,
$O(n, R)$	$\{X \in GL(n, R) : X^t = X^{-1}\}$,
$SO(n, R)$	$O(n, R) \cap SL(n, R)$,
$O(p, q)$	$\{X \in GL(p+q, R) : I_{p,q} X^t I_{p,q}^{-1} = X^{-1}\}$,
$SO(p, q)$	$O(p, q) \cap SL(p+q, R)$,
$Sp(n, C)$	$\{X \in GL(2n, C) : J_n X^t J_n^{-1} = X^{-1}\}$,
$Sp(n, R)$	$Sp(n, C) \cap GL(n, R)$,
$U(n)$	$\{X \in GL(n, C) : \bar{X}^t = X^{-1}\}$,
$SU(n)$	$U(n) \cap SL(n, C)$.

TABLE 2.2

LIE ALGEBRAS

$gl(n, C)$	$n \times n$ complex matrices,
$gl(n, R)$	$n \times n$ real matrices,
$sl(n, C)$	$\{X \in gl(n, C) : \mathrm{tr}(X) = 0\}$,
$sl(n, R)$	$sl(n, C) \cap gl(n, R)$,
$so(n, R)$	$\{X \in gl(n, R) : X^t = -X\}$,
$so(p, q)$	$\{X \in gl(p+q, R) : I_{p,q} X^t I_{p,q}^{-1} = -X\}$,
$sp(n, C)$	$\{X \in gl(2n, C) : J_n X^t J_n^{-1} = -X\}$,
$sp(n, R)$	$sp(n, C) \cap gl(2n, R)$,
$u(n)$	$\{X \in gl(n, C) : \bar{X}^t = -X\}$,
$su(n)$	$u(n) \cap sl(n, C)$.

For more details on matrix groups the reader should consider the work of Chevalley [1946, Chapter 1] and Helgason [1962, p. 339].

4. Tangents and Cotangents

Let $M \subset R^3$ be a well-behaved surface given by the differentiable function $z = f(x, y)$ and going through the point $p = (0, 0, 0)$. Then from calculus the tangent plane to M at p is given by the equation

$$z = x\, \partial f(0, 0)/\partial x + y\, \partial f(0, 0)/\partial y \qquad \text{for} \quad x, y \in R.$$

If this plane is cut by the plane $x = 0$, then the equation of the line of intersection is $z = y\,\partial f(0, 0)/\partial y$ and we obtain the vector $(0, 1, \partial f(0, 0)/\partial y)$ in the tangent plane. Similarly $(1, 0, \partial f(0, 0)/\partial x)$ is in the tangent plane. These two vectors which give the tangent plane are determined by the partial differentiation of f. Thus we are led to study operators on real-valued functions which have the properties of differentiation and we now abstract this situation to manifolds.

First recall that for a C^∞-manifold M and for $p \in M$ the set $F(p) = C^\infty(p)$ of C^∞-functions at $p \in M$ is an associative algebra using the pointwise operations: Let U, V be open sets of M containing p and let $f: U \to R$, $g: V \to R$ be in $F(p)$. Then define for $a, b \in R$

$$af + bg : U \cap V \to R : q \to af(q) + bg(q) \quad \text{and} \quad fg : U \cap V \to R : q \to f(q)g(q).$$

Definition 2.16 A **tangent at** $p \in M$ is a mapping $L : F(p) \to R$ such that for all $f, g \in F(p)$ and $a, b \in R$,

(a) $L(af + bg) = aL(f) + bL(g)$;
(b) $L(fg) = L(f)g(p) + f(p)L(g)$.

That is, L is a **derivation** of $F(p)$ into R. Let $T_p(M)$, or $T(M, p)$, or M_p denote the set of tangents at $p \in M$.

Example (1) For $p \in M = R^m$ and for fixed $X \in R^m$ the map

$$L_X : F(p) \to R : f \to [Df(p)](X)$$

is a tangent at p using Proposition 1.3 concerning the product rule.

Proposition 2.17 Let M be an m-dimensional C^∞-manifold and let $p \in M$.

(a) If $f, g \in F(p)$ and $f(q) = g(q)$ for all q in a neighborhood U of p, then $L(f) = L(g)$ for all $L \in T(M, p)$.

(b) $T(M, p)$ is a vector space over R.

PROOF (a) The function k defined on U by $k(q) = 1$ for all $q \in U$ is in $F(p)$ and we have for any $L \in T(M, p)$ that

$$\begin{aligned} L(k) &= L(k^2), \qquad \text{using} \quad 1 = 1^2 \\ &= L(k)k(p) + k(p)L(k) \\ &= 2L(k). \end{aligned}$$

Thus $L(k) = 0$. Now we have $f = kf = kg$ on U and therefore

$$\begin{aligned} L(f) &= L(k)f(p) + k(p)L(f) \\ &= L(kf) = L(kg) \\ &= L(k)g(p) + k(p)L(g) = L(g). \end{aligned}$$

(Can the "bump function" of exercise 6, Section 1.4 be used above?)

(b) For L_1 and L_2 in $T(M, p)$ and for $a, b \in R$ we see that $aL_1 + bL_2$ is a linear operator; that is, it satisfies (a) of the definition. Also for $f, g \in F(p)$ we have

$$\begin{aligned}(aL_1 + bL_2)(fg) &= aL_1(fg) + bL_2(fg)\\ &= a\big(L_1(f)g(p) + f(p)L_1(g)\big) + b\big(L_2(f)g(p) + f(p)L_2(g)\big)\\ &= (aL_1 + bL_2)(f)g(p) + f(p)(aL_1 + bL_2)(g).\end{aligned}$$

Thus $aL_1 + bL_2 \in T(M, p)$.

We shall now show that the vector space dimension of $T(M, p)$ is m; that is, equal to the manifold dimension of M. We shall do this by taking a chart (U, x) of M at p such that for $u_i : x(U) \to R$ where the u_i are coordinate functions of R^m, the partial derivative operators $D_i(x(p)) = \partial/\partial u_i(x(p))$ in R^m for $i = 1, \ldots, m$ eventually yield a basis $\partial_i(p)$ for $i = 1, \ldots, m$ of $T(M, p)$.

Thus let (U, x) be a fixed chart at p in M and let $f \in F(p)$, where f is defined on an open neighborhood V of p with $f\colon V \to R$ of class C^∞. Now f is of class C^∞ on the neighborhood $U \cap V \subset U$ so that we can write f in terms of the fixed coordinates $x = (x_1, \ldots, x_m)$ where $x_i = u_i \circ x$. Therefore for $D = x(U)$ an open set in R^m the function $g = f \circ x^{-1} : D \to R$ is C^∞ on D. Thus $f = g \circ x = g(x_1, \ldots, x_m)$ where we write $g = g(u_1, \ldots, u_m)$ on D. We now define the maps

$$\partial_i : F(p) \to F(p) : f \to \partial(f \circ x^{-1})/\partial u_i \circ x$$

which are called **coordinate vector fields** relative to (U, x); that is, we form the real-valued C^∞-function $h = \partial(f \circ x^{-1})/\partial u_i = \partial g/\partial u_i$ defined on D to obtain the function $h \circ x$ which is in $F(p)$. Sometimes the notations

$$\partial_i = \partial/\partial x_i \qquad \text{and} \qquad \partial_i f = \partial f/\partial x_i$$

are used. The mapping ∂_i has the following properties for $f, g \in F(p)$ and $a, b \in R$,

(1) $\partial_i(af + bg) = a\,\partial_i f + b\,\partial_i g$;

(2) $\partial_i(fg) = (\partial_i f)g + f(\partial_i g)$.

Note that for the coordinate functions $x_j = u_j \circ x$ on U we have from the above definition

$$\partial_i x_j = \partial u_j/\partial u_i = \delta_{ij}$$

and for the constant function $f(q) = c$ for $q \in U$, we have $\partial_i f = 0$.

Next we define an element $\partial_i(p) \in T(M, p)$ as follows: For $f \in F(p)$ we obtain $\partial_i f \in F(p)$, then evaluate $(\partial_i f)(p) \in R$. Thus

$$\partial_i(p)f = (\partial_i f)(p)$$

and from (1), (2) above we see $\partial_i(p)$ satisfies the definition of a tangent at p.

Proposition 2.18 Let (U, x) be a fixed chart at $p \in M$ where $x = (x_1, \ldots, x_m)$. Then the vector space $T(M, p)$ has basis

$$\partial_1(p), \ldots, \partial_m(p)$$

and any $L \in T(M, p)$ has the unique representation

$$L = \sum_{i=1}^{m} a_i \partial_i(p),$$

where $a_i = L(x_i) \in R$. Thus the manifold dimension of M equals the vector space dimension of $T(M, p)$.

PROOF We can assume that $x(p) = 0$ since a translation $x_i = y_i + t$ yields $\partial/\partial x_i = \partial/\partial y_i$. Now from Section 1.4 it is easy to see that any real-valued C^∞-function g defined on $D = x(U)$ has the Taylor's formula expansion about the point $\theta = (0, \ldots, 0) \in D \subset R^m$

$$g = g(\theta) + \sum_{i=1}^{m} u_i g_i,$$

where u_i are coordinates on R^m and g_i are C^∞ at $\theta \in D$. Thus for the real valued function $f = g \circ x \in F(p)$ as previously discussed we obtain on U

$$\begin{aligned} f = g \circ x &= g(\theta) + \sum (u_j \circ x)(g_j \circ x) \\ &= g(\theta) + \sum x_j f_j, \end{aligned}$$

where $f_j = g_j \circ x \in F(p)$. We apply $\partial_i(p)$ to this equation

$$\begin{aligned} \partial_i(p) f &= (\partial_i f)(p) \\ &= 0 + \sum \partial_i(x_j f_j)(p) \\ &= \sum [(\partial_i x_j)(p) f_j(p) + x_j(p)\, \partial_i f_j(p)] = f_i(p) \end{aligned}$$

using $\partial_i x_j = \delta_{ij}$ and $x_j(p) = 0$. Next we apply L to the same equation

$$\begin{aligned} L(f) &= L\big(g(\theta)\big) + \sum L(x_j f_j) \\ &= 0 + \sum [(L x_j) f_j(p) + x_j(p) L(f_j)] = \sum a_j \partial_j(p) f \end{aligned}$$

using the preceding equation. Thus $L = \sum a_j \partial_j(p)$. The elements $\partial_1(p), \ldots, \partial_m(p)$ in $T(M, p)$ are linearly independent. For if $\sum a_j \partial_j(p) = 0$, then applying to coordinate functions,

$$0 = 0(x_i) = \sum a_j \partial_j(p)(x_i) = a_i$$

using $\partial_j(p)(x_i) = \partial_j x_i(p) = \delta_{ij}$.

REMARK Let (U, x) and (V, y) be charts at $p \in M$. Then on $U \cap V$ we have the coordinate functions x_i and y_j defined. Then $\partial/\partial x_i(p)$ and $\partial/\partial y_j(p) = L$ are in $T(M, p)$ and according to Proposition 2.18 we can represent L by

$$\partial/\partial y_j(p) = \sum_{i=1}^{m} \partial x_i/\partial y_j(p)\, \partial/\partial x_i(p),$$

where the matrix $((\partial x_i/\partial y_k)(p))$ is the nonsingular Jacobian matrix obtain by writing $x_i = x_i(y_1, \ldots, y_m)$. Thus we have the matrix for the change of basis in $T(M, p)$ when we change charts at $p \in M$.

Examples (2) For $p \in M = R^m$ and $X \in R^m$, let $L_X \in T(M, p)$ be the tangent given in example (1) of this section. Then the map $R^m \to T(M, p)$: $X \to L_X$ is linear because $L_{aX+bY}(f) = Df(p)(aX + b\,Y) = (aL_X + bL_Y)(f)$. Also this map is an isomorphism. (Why?) Thus at each point $p \in M = R^m$ we can attach the tangent space which is isomorphic to M itself.

(3) Let N be a group with identity e and let $(x, y) = xyx^{-1}y^{-1}$ be in N and for A, B subsets of N let (A, B) be the subgroup generated by all commutators (x, y) with $x \in A$, $y \in B$. Then for $N_1 = (N, N)$, $N_{k+1} = (N_k, N)$ we have

$$N \supset N_1 \supset \cdots \supset N_k \supset \cdots$$

and call N **nilpotent** if there exists k with $N_k = \{e\}$. Now let N be the submanifold of $SL(V)$ consisting of the nilpotent subgroup given by the set of triangular matrices

$$\begin{bmatrix} 1 & & & * \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

where $*$ denotes arbitrary elements from R. We shall now show that the vector space $T_I(N)$ is isomorphic to the vector space of all triangular (nilpotent) matrices

$$\begin{bmatrix} 0 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}$$

and denote this vector space of matrices by n. To show the isomorphism we shall use the exp mapping by showing $\exp : n \to N$ is locally invertible. Now for $A \in n$ we see that the associative products $A^2, A^3, \ldots, A^k$ are all in n and since A is a nilpotent matrix $A^m = 0$. Thus we see that

$$\exp A = I + A + \cdots + A^{m-1}/(m-1)!$$

is in N so that $\exp : n \to N$. Also $D\exp(0)$ is invertible. Thus as before, there exist a neighborhood U_0 of 0 in n and a neighborhood of U_I of I in N so that $(U_I, \log)$ is a chart in N at I. Consequently we have, since U_0 is open in n,

$$\dim U_I = \dim U_0 = \dim n.$$

However $\dim T_I(N) = \dim U_I$ and since the vector spaces $T_I(N)$ and n have the same dimension, they are isomorphic.

Exercises (1) Show that n is a **nilpotent** Lie algebra. Thus first show $[n, n] = \{\sum [A_i, B_i] : A_i, B_i \in n\} \subset n$. Next define

$$n^1 = [n, n] \qquad \text{and} \qquad n^{k+1} = [n^k, n]$$

and note that $n \supset n^1 \supset \cdots \supset n^k \supset \cdots$. So finally show $n^p = 0$ for some p. This will show that the nilpotent (Lie) group N is such that the tangent space $T_I(N)$ is vector space isomorphic to a nilpotent Lie algebra n.

(2) Let M, N be C^∞-manifolds and let $p \in M$, $q \in N$. Then show

$$T(M \times N, (p, q)) \cong T(M, p) \times T(N, q) \cong T_p(M) \oplus T_q(N).$$

Recall that if V is an m-dimensional vector space over R, then its **dual space** $V^* = \mathrm{Hom}(V, R)$. Elements of V^* are called **linear functionals** and the map

$$V \times V^* \to R : (X, f) \to f(X)$$

is bilinear and is frequently written

$$f(X) = \langle X, f \rangle.$$

Now for any basis $X_1, \ldots, X_m$ of V we have the **dual basis** $f_1^*, \ldots, f_m^*$ given by

$$f_j^*(X_i) = \langle X_i, f_j^* \rangle = \delta_{ij}.$$

From this we see any $X \in V$ can be written in the form

$$X = \sum f_i^*(X) X_i.$$

Definition 2.19 The **cotangent space** at $p \in M$ is the dual space of $T_p(M)$ and is denoted by $T_p^*(M)$, or $T^*(M, p)$, or M_p^*. The elements of $T^*(M, p)$ are frequently called **differentials** at p and $T^*(M, p)$ is also called the **space of differentials** at p.

Now let $f \in F(p)$ and define the element $df \in T^*(M, p)$ by

$$df : T(M, p) \to R : L \to L(f);$$

that is, $df(L) = \langle L, df \rangle = L(f)$. Sometimes a more specific notation $df(p)$ or df_p will be used. In particular if (U, x) is a chart with x_i the coordinate functions, then a basis for $T(M, p)$ is given by $\partial_1(p), \ldots, \partial_m(p)$ and a dual basis for $T^*(M, p)$ is given by $dx_1(p), \ldots, dx_m(p)$ because they satisfy

$$\langle \partial_i(p), dx_j(p) \rangle = \partial(x_j)/\partial x_i(p) = \delta_{ij}.$$

Now for any $L \in T(M, p)$ and any $f \in F(p)$ we have from Proposition 2.18 that $L = \sum L(x_i)\, \partial_i(p)$ and therefore

$$\begin{aligned} df(L) = L(f) &= \sum L(x_i)(\partial_i f)(p) \\ &= \sum (\partial_i f)(p) L(x_i) \\ &= \sum (\partial_i f)(p)\, dx_i(L); \end{aligned}$$

that is,

$$df(p) = \sum (\partial_i f)(p)\, dx_i(p).$$

Combining various facts we have the following result.

Proposition 2.20 Let M be an m-dimensional manifold and let $f_1, \ldots, f_r \in F(p)$ for $p \in M$.

(a) Each $f \in F(p)$ equals $g(f_1, \ldots, f_r)$ on a suitable neighborhood $V = V(f)$ of p, where $g : R^r \to R$ is of class C^∞ if and only if $df_1(p), \ldots, df_r(p)$ generate the cotangent space $T^*(M, p)$.

(b) The functions $f_1, \ldots, f_m$ (that is, $r = m$) are the coordinates of some chart (U, f) at p where $f = (f_1, \ldots, f_m)$ if and only if the set $df_1(p), \ldots, df_m(p)$ is a basis of $T^*(M, p)$.

PROOF (a) Let (U, x) be a chart at p and suppose each $f \in F(p)$ equals $g(f_1, \ldots, f_r)$ on $V \cap U$. Then each of the coordinate functions

$$x_i = g_i(f_1, \ldots, f_r)$$

and therefore $dx_i = \sum \partial_k g_i(p)\, df_k$. However, since the dx's generate $T^*(M, p)$, the df's also generate $T^*(M, p)$. Conversely, assume the df's generate $T^*(M, p)$ and represent f_i in coordinates

$$f_i = h_i(x_1, \ldots, x_m).$$

Then we obtain

$$df_i = \sum \partial_k h_i(p)\, dx_k,$$

$i = 1, \ldots, r$. Now since the df's generate $T^*(M, p)$ the $m \times r$ matrix $(\partial_k h_i(p))$ has rank $m \leq r$. Thus we can assume that there exists a system of m functions

$$f_{i_j} = h_{i_j}(x_1, \ldots, x_m) \qquad \text{for} \quad j = 1, \ldots, m$$

which define a function $F: R^m \to R^m : x = (x_1, \ldots, x_m) \to (F_1(x), \ldots, F_m(x))$ where $F_j(x) = f_{i_j}(x)$. Also $DF(x(p))$ is invertible so that by the inverse function theorem we can write locally

$$x_i = k_i(f_{i_1}, \ldots, f_{i_m}),$$

where $k_i : R^m \to R$ are C^∞. However, each $f \in F(p)$ equals $G(x_1, \ldots, x_m)$ locally, where G is C^∞, and using the above expression for x's in terms of f_j's we have the results.

To show (b) just note that for $r = m$ we have $df_1, \ldots, df_m$ generate $T_p{}^*(M)$ if and only if they form a basis. Then we can use the above equations expressing $x_i = k_i(f_1, \ldots, f_m)$ and $f_j = h_j(x_1, \ldots, x_m)$ to see $f_1, \ldots, f_m$ are coordinates for some chart at $p \in M$.

Exercise (3) Let U be open in R^m and let $f: U \to R$ be of class C^∞. Compare $Df(p)$ and $df(p)$ for $p \in U$.

5. Tangent Maps (Differentials)

In the preceding section we considered a C^∞-map g from the manifold M into the manifold R and noted that the differential $df(p)$ is a linear map from the tangent space $T(M, p)$ into the vector space $R \cong T(R, f(p))$; this isomorphism uses example (2) of Section 2.4. We shall generalize this situation by showing that a C^∞-map $f: M \to N$ between two manifolds induces a linear map $df(p) : T(M, p) \to T(N, f(p))$. However, by means of coordinate functions this generalized situation reduces to that of the preceding section.

Definition 2.21 Let M and N be C^∞-manifolds and let $f: M \to N$ be a C^∞-mapping. The **differential of** f at $p \in M$ is the map

$$df(p) : T(M, p) \to T(N, f(p))$$

given as follows. For $L \in T(M, p)$ and for $g \in F(f(p))$, we define the action of $df(p)(L)$ on g by

$$[df(p)(L)](g) = L(g \circ f).$$

REMARKS (1) We shall frequently use the less specific notation df for $df(p)$ when there should be no confusion. Also we shall use the notation

$$Tf = Tf(p) = df(p)$$

and also call $Tf(p)$ the **tangent map** of f at p. This notation is very useful in discussing certain functors on categories involving manifolds.

(2) We note that for $g \in F(f(p))$ the function $g \circ f$ is in $F(p)$ so the operation $L(g \circ f)$ is defined. We must next show $Tf(L)$ is actually in $T(N, f(p))$ by showing it is a derivation. Thus for $g, h \in C^\infty(f(p))$,

$$\begin{aligned} Tf(L)(ag + bh) &= L(a(g \circ f) + b(h \circ f)) \\ &= aL(g \circ f) + bL(h \circ f) \\ &= a[Tf(L)](g) + b[Tf(L)](h) \end{aligned}$$

and the product rule is also easy.

The following result shows that $df(p)$ is the correct generalization for $Df(p)$ of Section 1.2, where $f: U \to W$ is a C^∞-map of an open set U in R^m and W is some Euclidean space.

Proposition 2.22 Let $f: M \to N$ be a C^∞-map of C^∞-manifolds and let $p \in M$. Then the map

$$Tf(p): T(M, p) \to T(N, f(p))$$

is a linear transformation; that is, $Tf(p) \in \operatorname{Hom}(T(M, p), T(N, f(p)))$. Furthermore if (U, x) is a chart at p and (V, y) is a chart at $f(p)$, then $Tf(p)$ has a matrix which is the Jacobian matrix of f represented in these coordinates.

PROOF Let $X, Y \in T(M, p)$. Then for $a, b \in R$ and $g \in F(f(p))$ we have

$$\begin{aligned} [Tf(aX + bY)](g) &= (aX + bY)(g \circ f) \\ &= aX(g \circ f) + bY(g \circ f) \\ &= [a\, Tf(X) + b\, Tf(Y)](g) \end{aligned}$$

so that $Tf(aX + bY) = a\, Tf(X) + b\, Tf(Y)$. Next let $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_n)$ be the given coordinate functions so that we can represent f in terms of coordinates in the neighborhood V by

$$f_k = y_k \circ f = f_k(x_1, \ldots, x_m) \qquad \text{for} \quad k = 1, \ldots, n.$$

Now let $\partial/\partial x_i = \partial_i(p)$ and $\partial/\partial y_i = \partial_i(f(p))$ determine a basis for $T(M, p)$ and $T(N, f(p))$, respectively. Thus to determine a matrix for Tf we compute its action on the basis $\partial/\partial x_i$ in $T(M, p)$. Let

$$Tf(\partial/\partial x_i) = \sum_j b_{ji}\, \partial/\partial y_j$$

be in $T(N, f(p))$. Then we evaluate the matrix (b_{ji}) using the fact that $y_k \in F(f(p))$ as follows

$$\begin{aligned} \partial f_k/\partial x_i(p) &= \partial_i(p)(y_k \circ f) \\ &= [Tf(\partial_i(p))](y_k) \\ &= \sum_j b_{ji}\, \partial(y_k)/\partial y_j = b_{ki} \end{aligned}$$

using $\partial(y_k)/\partial y_j = \delta_{kj}$. Thus $(b_{ji}) = (\partial f_j/\partial x_i(p))$ is the desired Jacobian matrix.

Proposition 2.23 (Chain rule) Let M, N, and P be C^∞-manifolds and let

$$f : M \to N \qquad \text{and} \qquad g : N \to P$$

be C^∞-maps. Then for $p \in M$,

$$T(g \circ f)(p) = Tg(f(p)) \cdot Tf(p)$$

which is a composition of homomorphisms of tangent spaces.

PROOF Let $X \in T(M, p)$ and $h \in F((g \circ f)(p))$. Then using definitions and $df(p)X \in T(N, f(p))$ we have

$$\begin{aligned}[T(g \circ f)(p)X](h) &= X((h \circ g) \circ f) \\ &= [df(p)X](h \circ g) \\ &= [dg(f(p))\,(df(p))X](h) \\ &= [Tg(f(p)) \cdot Tf(p)X](h).\end{aligned}$$

REMARKS (3) If U is open in a C^∞-manifold M, then U is a C^∞-submanifold such that the inclusion map $i : U \to M : x \to x$ is C^∞. Also for $u \in U$, $Ti(u) : T(U, u) \to T(M, u)$ is an isomorphism and we identify these tangent spaces by this isomorphism.

Many of the preceding results on submanifolds can be easily expressed in terms of tangent maps and are usually taken as definitions. Thus let M and N be C^∞-manifolds of dimension m and n, respectively, and let

$$f : M \to N$$

be a C^∞-map. Then we have the following results.

The **inverse function theorem** can be stated as follows: If $p \in M$ is such that

$$Tf(p) : T(M, p) \to T(N, f(p))$$

is an isomorphism, then $m = n$ and f is a local diffeomorphism. Thus there exists a neighborhood U of p in M such that

(1) $f(U)$ is open in N;
(2) $f : U \to f(U)$ is injective;
(3) the inverse map $f^{-1} : f(U) \to U$ is C^∞.

We now consider separately the injective and surjective parts of the above homomorphism $Tf(p)$; this was discussed in Section 2.3.

We have f is an **immersion** if and only if $Tf(p)$ is injective for all $p \in M$. In case f is injective, f is an **embedding**. Also $f(M)$ is a **submanifold** of N if f is an embedding and if $f(M)$ has a C^∞-structure such that $f : M \to f(M)$ is a diffeomorphism. Thus from preceding results we have that the following are equivalent:

(1) $Tf(p)$ is injective;

(2) there exists a chart (U, x) at p in M and a chart (V, y) at $f(p)$ in N such that $m \leq n$ and $x_i = y_i \circ f$ for $i = 1, \ldots, m$ and $y_j \circ f = 0$ for $j = m + 1, \ldots, n$;

(3) there exists a neighborhood U of p in M and a neighborhood V of $f(p)$ in N and there exists a C^∞-map $g : V \to U$ such that $f(U) \subset V$ and $g \circ f$ is the identity $| U$.

A C^∞-function $f: M \to N$ satisfying (1) at $p \in M$ is called **regular** at p. If f is regular at every $p \in M$, then it is also called a **regular function.**

We have f is a **submersion** if and only if $Tf(p)$ is surjective for all $p \in M$. Also the following are equivalent:

(1) $Tf(p)$ is surjective;

(2) there exists a chart (U, x) at p in M and a chart (V, y) at $f(p)$ in N such that $m \geq n$ and $x_i = y_i \circ f$ for $i = 1, \ldots, n$;

(3) there exists a neighborhood U of p in M and a neighborhood V of $f(p)$ in N and a C^∞-map $g : V \to U$ such that $f(U) \supset V$ and $f \circ g$ is the identity $| V$.

Using the surjective nature of $Tf(p)$ we reformulate Proposition 2.13 and construct submanifolds using the following version of the **implicit function theorem**: Let $f: M \to N$ be a C^∞-map of C^∞-manifolds and let $m = \dim M \geq \dim N = n$. Let $q \in f(M)$ be a fixed element and let

$$f^{-1}(q) = \{p \in M : f(p) = q\}.$$

If for each $p \in f^{-1}(q)$ we have $Tf(p) : T(M, p) \to T(N, f(p))$ is surjective, then $f^{-1}(q)$ has a manifold structure for which the inclusion map $i : f^{-1}(q) \to M$ is C^∞. Thus $f^{-1}(q)$ is a submanifold of M. Furthermore the underlying topology of the submanifold $f^{-1}(q)$ is the relative topology and the dimension of $f^{-1}(q)$ is $m - n$.

Examples (1) We next consider the special case of $f: M \to N$ where $M = R$ or $N = R$. First let $N = R$; that is, $f \in C^\infty(M)$. Then combining the notation of Sections 2.4 and 2.5 we have $Tf(p) = df(p)$ and for $X \in T(M, p)$ we have $Tf(p)(X) \in T(R, f(p)) \cong R$. Thus for $u : R \to R : t \to t$ the coordinate function on the manifold R, we have for some $a \in R$

$$Tf(p)X = a(d/du)$$

and as before $a = a(d(u)/du) = [Tf(p)X](u) = X(u \circ f) = X(f)$, using $u(t) = t$. Consequently the map

$$T(R, f(p)) \to R : a(d/du) \to a$$

is the isomorphism which yields

$$Tf(p) : T(M, p) \to R;$$

that is, which yields the cotangent space.

(2) Next consider $f: R \to M$ formulated in terms of curves. Thus let $I = (a, b)$ and let $\alpha : I \to M$ be a C^∞-curve which admits an extension $\tilde{\alpha} : (a - \varepsilon, b + \varepsilon) \to M$ (Definition 2.14). The **tangent vector** to α at $t \in I$ is denoted by $\dot{\alpha}(t)$ and defined by

$$\dot{\alpha}(t) = [T\alpha(t)](d/du),$$

where $u : R \to R$ is the coordinate function discussed above.

Now let $X \in T(M, p)$. Then there exists a curve $\alpha : I \to M$, where I is an interval containing $0 \in R$ such that $\alpha(0) = p$ and $\dot{\alpha}(0) = X$, for let (U, x) be a coordinate system at p with $x(p) = 0$ and find a curve $\beta : R \to x(U) \subset R^m$ with $\beta(0) = 0$ and $\dot{\beta}(0) = [Tx(p)](X)$; that is, β a straight line. Then $\alpha = x^{-1} \circ \beta$ is the desired curve

$$\begin{aligned}\dot{\alpha}(0) &= [T(x^{-1} \circ \beta)(0)](d/du)\\ &= Tx^{-1}(x(p)) \cdot \dot{\beta}(0), \qquad \text{using the chain rule and} \quad x(p) = 0\\ &= Tx^{-1}(x(p)) \cdot [Tx(p)](X) = X\end{aligned}$$

and

$$\alpha(0) = x^{-1}(\beta(0)) = p.$$

Also for $f \in F(p)$ we have

$$\begin{aligned}X(f) &= \dot{\alpha}(0)(f)\\ &= [(T\alpha(0))(d/du)](f) = d/du(0)(f \circ \alpha).\end{aligned}$$

Let (U, x) be a chart on M and let $\alpha : (a, b) \to U \subset M$ be a C^∞-curve as above. Then for $t \in (a, b)$ we can represent

$$\dot{\alpha}(t) = [T\alpha(t)](d/du) = \sum a_k\, \partial_k(\alpha(t)) \in T(M, \alpha(t))$$

and evaluate the coefficients $a_i = a_i(t)$ using the dual basis of differentials as follows.

$$\begin{aligned}a_i &= dx_i(\textstyle\sum a_k\, \partial_k)\\ &= dx_i(\dot{\alpha})\\ &= dx_i[d\alpha(d/du)], \qquad \text{notation}\\ &= d(x_i \circ \alpha)(d/du), \qquad \text{chain rule}\\ &= d/du(x_i \circ \alpha),\end{aligned}$$

where we use the definition of differential of a function applied to a tangent (note paragraph following Definition 2.19). Thus as in calculus the tangent vector to a curve α is obtained by differentiating its coordinate representation.

(3) Consider the special case when $M = G = GL(n, R)$. We shall construct an explicit vector space isomorphism of $g = gl(n, R)$ onto $T_I(G)$. Thus for any fixed $X \in g$ let

$$\alpha : R \to G : t \to \exp tX.$$

Then $\alpha = \tilde{\alpha}$ and define an element $\bar{X} \in T_I(G)$ by

$$\bar{X}(f) = \dot{\alpha}(0)(f)$$

for any $f \in F(I)$. From the preceding example we actually have $\bar{X} \in T_I(G)$ since $\alpha(0) = I$. Next note that

$$\begin{aligned} \bar{X}(f) &= [(T\alpha(0))(d/du)](f) \\ &= d/du(0)(f \circ \alpha) \\ &= \lim_{t \to 0} [f(\exp tX) - f(I)]/t \\ &= [Df(I)](X). \end{aligned} \tag{$*$}$$

Now define the mapping

$$\varphi : g \to T_I(G) : X \to \bar{X},$$

where φ is well defined and for X, $Y \in g$ and $a, b \in R$ we use Eq. $(*)$ to obtain, for any $f \in F(I)$,

$$\begin{aligned} \varphi(aX + bY)(f) &= \overline{aX + bY}(f) \\ &= [Df(I)](aX + bY) \\ &= a\,Df(I)(X) + b\,Df(I)(Y) \\ &= a\bar{X}(f) + b\bar{Y}(f) \\ &= [a\varphi(X) + b\varphi(Y)](f) \end{aligned}$$

so that φ is a vector space homomorphism. Next suppose $\bar{X} = 0$ and let $u_1, \ldots, u_m$ $(m = n^2)$ be coordinates in $g\,(\cong R^m)$ corresponding to a basis $X_1, \ldots, X_m$ of g. Let $X = \sum x_i X_i \in g$ with $\varphi(X) = \bar{X} = 0$ and let $f_i = u_i \circ \log \in F(I)$ as previously discussed. Then $f_i(I) = 0$ and since $\bar{X} = 0$,

$$\begin{aligned} 0 &= \bar{X} f_i \\ &= \lim_{t \to 0} [f_i(\exp tX) - f_i(I)]/t, \qquad \text{Eq. } (*) \\ &= x_i \end{aligned}$$

so that $X = 0$ and φ is an isomorphism. We frequently omit this isomorphism and just use the most convenient identification for a given problem.

(4) Let $f : G \to G$ be a C^∞-automorphism of $G = GL(V)$. Then from

Section 1.6 we see that the "tangent map" $Df(I)$ is an automorphism of the Lie algebra $g = gl(V)$. Thus relations on the Lie group are translated to relations on the Lie algebra by the tangent map.

6. Tangent Bundle

In this section we shall show how to make the collection of tangent spaces of a manifold into a manifold. We also discuss mappings of such manifolds and use them to define vector fields in the next section.

Definition 2.24 Let M be a C^∞-manifold of dimension m and let

$$T(M) = \bigcup\{T(M, p) : p \in M\}$$

which is a disjoint union. We call $T(M)$ the **tangent bundle** of M.

We now make $T(M)$ into a manifold (Fig. 2.10). We shall frequently denote the points of $T(M)$ by the pairs (p, Y) where $p \in M$ and $Y \in T(M, p)$; the p is unnecessary in this notation but convenient. First $T(M)$ is a Hausdorff space as follows. Let

$$\pi : T(M) \to M : (p, Y) \to p$$

be the **projection map**. For $(p, Y) \in T(M)$ let (U, x) be a chart at p in the atlas $\mathscr{A}$ of M. Then $\pi^{-1}(U) = \{(q, X) \in T(M) : q \in U\}$. Now if $(q, X) \in \pi^{-1}(U)$, then in terms of coordinates $x(q) = (x_1(q), \ldots, x_m(q))$ and $X = \sum a_j\, \partial/\partial x_j(q)$, where $a_j = a_j(q)$. The map

$$\phi_U : \pi^{-1}(U) \to R^{2m} : (q, X) \to (x_1(q), \ldots, x_m(q), a_1, \ldots, a_m)$$

is injective and there is a unique topology on $T(M)$ such that for all $(U, x) \in \mathscr{A}$, the maps ϕ_U are homeomorphisms (why?). This topology defined by the sets $\pi^{-1}(U)$ can easily be seen to be Hausdorff using the fact that M and R^m are Hausdorff. Also note that since M has a countable basis of neighborhoods, then so does $T(M)$.

Next we define a C^∞-atlas on $T(M)$ so that the projection map $\pi : T(M) \to M$ is a C^∞-map. Thus for each $(p, Y) \in T(M)$ let $\pi^{-1}(U)$ be a neighborhood of (p, Y) where (U, x) is a chart at p and let $\phi(U) \equiv \phi_U$: $\pi^{-1}(U) \to R^{2m}$ be the above homeomorphism. We claim that $(\pi^{-1}(U), \phi(U))$ is a chart at (p, Y). Thus we must show any two such coordinate neighborhoods are compatible. Therefore, let (U, x), (V, y) be charts at p where the x and y

are C^∞-related by $x = f(y)$ on $U \cap V$; that is, $f = x \circ y^{-1}$. Now in terms of coordinates let $x_k = f_k(y_1, \ldots, y_m)$,

$$\phi(U) = (z_1, \ldots, z_m, z_{m+1}, \ldots, z_{2m}), \qquad \phi(V) = (w_1, \ldots, w_m, w_{m+1}, \ldots, w_{2m})$$

where

$$z_i(q, X) = x_i(q) \qquad \text{for} \quad i = 1, \ldots, m,$$
$$z_{j+m}(q, X) = a_j \qquad \text{for} \quad j = 1, \ldots, m,$$

and similarly $w_i(q, X) = y_i(q)$ for $i = 1, \ldots, m$ and $w_{i+m}(q, X) = b_i$ where $X = \sum b_i \, \partial/\partial y_i(q)$.

Now for $(q, X) \in \pi^{-1}(U) \cap \pi^{-1}(V)$ we have first for $i = 1, \ldots, m$

$$z_i(q, X) = x_i(q) = f_i(y_1(q), \ldots, y_m(q)) = f_i(w_1(q, X), \ldots, w_m(q, X))$$

so that the first m coordinate functions are C^∞-related. Next for $j = 1, \ldots, m$ and for $X = \sum a_i \, \partial/\partial x_i(q)$ we note that

$$z_{j+m}(q, X) = a_j(q) = X(x_j) = dx_j(X)$$

and similarly $dy_j(X) = w_{j+m}(q, X)$ for $j = 1, \ldots, m$. Thus by the transformation law for differentials (note remark following Proposition 2.18),

$$\begin{aligned} z_{j+m}(q, X) = dx_j(X) &= \sum_k \partial x_j/\partial y_k(q) \, dy_k(X) \\ &= \sum_k f_{jk}(y_1(q), \ldots, y_m(q)) \, dy_k(X) \\ &= \sum_k f_{jk}(w_1(q, X), \ldots, w_m(q, X)) w_{m+k}(q, X) \end{aligned}$$

which is a C^∞-relationship where $f_{ik} = \partial f_i/\partial y_k$ recalling $x_i = f_i(y_1, \ldots, y_m)$. Thus all the coordinates are compatible. The Hausdorff space $T(M)$ with the maximal atlas determined by the above charts is a C^∞-manifold and we shall always consider the tangent bundle with this C^∞-structure.

Finally we note that the projection map $\pi : T(M) \to M : (p, Y) \to p$ is C^∞. For let (U, x) be a chart at p and let $\phi(U) = (z_1, \ldots, z_{2m})$ be a coordinate system at (p, Y) as above. Then in terms of coordinates, $x_i \circ \pi(q, X) = z_i(q, X)$ for $i = 1, \ldots, m$ which shows that the coordinate expressions $x_i \circ \pi = z_i$ are C^∞ (see Fig. 2.10).

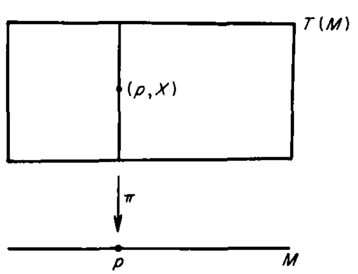


Fig. 2.10.

We now discuss real vector bundles [Lang, 1962; Loos, 1969].

Definition 2.25 A **vector bundle** E over a C^∞-manifold M is given by the following.

(a) E is a C^∞-manifold.
(b) There is a C^∞-surjection $\pi : E \to M$ called the **projection map**.
(c) Each **fiber** $E_p = \pi^{-1}(p)$ has the structure of a vector space over R.
(d) E is locally trivial; that is, there is a fixed integer n so that for each $p \in M$ there exists an open neighborhood U of p such that $U \times R^n$ is diffeomorphic to $\pi^{-1}(U)$ by a diffeomorphism ϕ so that the accompanying diagram is commutative, where pr_1 is the projection onto the first factor; specifically,

$$\begin{tikzcd} U \times R^n \arrow[rr, "\phi"] \arrow[dr, "pr_1"'] & & \pi^{-1}(U) \arrow[dl, "\pi"] \\ & U & \end{tikzcd}$$

$\pi \circ \phi(q, X) = q$. Furthermore we require for each $q \in U$ that $\phi(q, \quad)$ is a vector space isomorphism of R^n onto $E_q = \pi^{-1}(q)$.

Examples (1) $E = T(M)$ the tangent bundle where $n = m$ and $R^m \cong T(M, p) = E_p$.

(2) Let M be a C^∞-manifold and let

$$T^*(M) = \bigcup \{T^*(M, p)\} : p \in M\},$$

which is called the **cotangent bundle**. Then analogous to the construction of $T(M)$ we make $T^*(M)$ into a C^∞-manifold. Thus for $E = T^*(M)$ we see that $T^*(M)$ is a vector bundle and $E_p = \pi^{-1}(p) = T^*(M, p)$.

Definition 2.26 Let M and N be C^∞-manifolds and let E and E' be vector bundles over M and N, respectively. A **bundle homomorphism** is a pair of (surjective) maps (F, f) such that:

(a) $F : E \to E'$ and $f : M \to N$ are C^∞-maps;
(b) the accompanying diagram is commutative; that is, $\pi' \circ F = f \circ \pi$.

$$\begin{tikzcd} E \arrow[r, "F"] \arrow[d, "\pi"'] & E' \arrow[d, "\pi'"] \\ M \arrow[r, "f"'] & N \end{tikzcd}$$

Thus for each $p \in M$, $F(\pi^{-1}(p)) \subset (\pi')^{-1}(f(p))$.

(c) For each $p \in M$, the restriction $F : E_p \to E'_{f(p)}$ of the fibers is a linear transformation of the corresponding vector space structures. Also (F, f) is a **bundle isomorphism** of E onto E' if it is a bundle homomorphism such that the maps F and f are surjective diffeomorphisms. It is easy to see that in this case the pair (F^{-1}, f^{-1}) is a bundle isomorphism of E' onto E.

Examples (3) Let $g : M \to N$ be a C^∞-map. Then we define the map

$$T(g) : T(M) \to T(N) : (p, X) \to \big(g(p), [T(g)(p)](X)\big)$$

where $X \in T(M, p)$ and therefore $[T(g)(p)](X) \in T\big(N, g(p)\big)$. Then (Tg, g) is a bundle homomorphism of $T(M)$ into $T(N)$ because the diagram

$$\begin{tikzcd} T(M) \arrow[r, "Tg"] \arrow[d, "\pi"'] & T(N) \arrow[d, "\pi"] \\ M \arrow[r, "g"'] & N \end{tikzcd}$$

is commutative and $T(g)(p) : T(M, p) \to T\big(N, g(p)\big)$ is a vector space homomorphism.

Next note if we also have another C^∞-map $h : L \to M$ of manifolds, then $g \circ h : L \to N$ is a C^∞-map and

$$T(g \circ h) = T(g) \circ T(h) : T(L) \to T(N),$$

so that $\big(T(g \circ h),\ g \circ h\big)$ is a bundle homomorphism. Thus T can be regarded as a covariant functor from the category whose objects are manifolds and morphisms are C^∞-maps into the category whose objects are vector bundles and morphisms are bundle homomorphisms [Loos, 1969].

It will be easy to see later that if G is a Lie group, then the tangent bundle $T(G)$ is a Lie group and is isomorphic as a vector bundle and as a Lie group to the Lie group $g \times G$ (semi-direct product) where g is the Lie algebra of G. Thus the tangent bundles which we want to consider are of a relatively simple nature.

Exercise (1) Let M and N be C^∞-manifolds and let $M \times N$ be the corresponding product manifold. Show the tangent bundle $T(M \times N)$ is bundle isomorphic to $T(M) \times T(N)$.

7. Vector Fields

We have previously discussed the coordinate vector fields $\partial_i = \partial/\partial x_i$ and saw that they were functions defined on a neighborhood U of $p \in M$ which assigns to each $q \in U$ a tangent vector $\partial_i(q) \in T(M, q)$.

Definition 2.27 Let M be a C^∞-manifold and let $T(M)$ be the corresponding tangent bundle. A **vector field on a subset** $A \subset M$ is a map $X : A \to T(M)$ such that $\pi \circ X = \text{idy} \mid A$. Thus X assigns to each $p \in A$ a tangent vector $X(q)$, where $X(q) \in T(M, q)$, but such that $p = \text{idy}(p) = (\pi \circ X)(p) = \pi(X(q)) = q$. That is, the tangent vector assigned to p by X is actually in $T(M, p)$. Also X is a C^∞-**vector field on** A if A is open and if for each $f \in C^\infty(A)$ the function Xf is in $C^\infty(A)$ where we define Xf by the action of the corresponding tangent vector: $(Xf)(p) = [X(p)](f)$. Thus X is C^∞ on M if and only if $X : M \to T(M)$ is a C^∞-mapping of manifolds.

Example (1) Let $M = R^2$ and let $A = B(0, r)$ the open ball of radius r and center 0. Then with coordinates u_1, u_2 on A a C^∞-vector field X on A can be written

$$X = a_1(u_1, u_2)\, \partial/\partial u_1 + a_2(u_1, u_2)\, \partial/\partial u_2$$

where the a_1 and a_2 are C^∞-functions on A as shown below. Thus a C^∞-vector field is a well-behaved variable tangent vector.

REMARKS We now consider a vector field on M locally in terms of coordinates. Thus let (U, x) be a chart on M with U open in M, then we have the following.

(1) The coordinate vector fields $\partial_i = \partial/\partial x_i$ are C^∞-vector fields on U. This follows from the previous discussion: $\partial_i(p) \in T(M, p)$ so that $(\pi \circ \partial_i)(p) = p$. Next if $f \in C^\infty(U)$, then $g = f \circ x^{-1} : x(U) \to R$ is C^∞ on the open set $x(U) \subset R^m$. Also $\partial_i(f) = \partial g/\partial u_i \circ x$ is C^∞ on U where $u_1, \ldots, u_m$ are coordinates on R^m.

(2) If X is a C^∞-vector field on U, then there exist functions $a_i \in C^\infty(U)$ such that $X = \sum a_i \partial_i$ on U. Furthermore the $a_i = X(x_i)$. Thus the functions $a_i : U \to R$ exist because for each $q \in U$, the tangents $\partial_i(q)$, $i = 1, \ldots, m$, form a basis of $T(M, q)$ and $X(q) = \sum a_i(q)\, \partial_i(q)$ for some $a_i(q) \in R$. The a_i are C^∞ since $\partial_k(x_i) = \delta_{ki}$ and $a_i = \sum a_k \partial_k(x_i) = X(x_i)$ which is in $C^\infty(U)$. Also note that if X is a C^∞-vector field on M, then the restriction $X \mid U$ is a C^∞-vector field on U and has the above expression in coordinates. Thus summarizing

we see that a vector field X is C^∞ on M if and only if for every chart (U, x) the corresponding component functions $a_i = X(x_i)$ are in $C^\infty(U)$.

The following result is frequently taken as a definition [Helgason, 1962].

Proposition 2.28 We have that X is a C^∞-vector field on M if and only if X is a derivation of the algebra $C^\infty(M)$ into $C^\infty(M)$.

PROOF For each $f, g \in C^\infty(M)$ and $a, b \in R$ the properties

$$X(af + bg) = aX(f) + bX(g) \qquad \text{and} \qquad X(fg) = (Xf)g + f(Xg)$$

follow from the corresponding properties for tangents (Definition 2.16). Also by definition X is C^∞ if and only if $Xf \in C^\infty(M)$.

We have seen that a C^∞-vector field X on M restricts to a tangent $X(p)$ and we now consider the converse of extending a tangent to a vector field.

Proposition 2.29 Let M be a C^∞-manifold and let $X \in T(M, p)$. Then there exists a vector field $\tilde{X}$ which is C^∞ on M such that $\tilde{X}(p) = X$.

PROOF We can choose a chart (U, x) at p such that $X = \sum b_i\, \partial_i(p)$. Thus defining the constant functions $a_i : U \to R : q \to b_i$ we see that $Y = \sum a_i\, \partial_i$ is a C^∞-vector field on U such that $X = Y(p)$. Now let $\phi : M \to R$ be a C^∞ "bump function" at p; that is, from exercise (6), Section 1.4 we have $p \in D \subset U$, where D is an open neighborhood of p and the C^∞-function ϕ satisfies $0 \leq \phi(x) \leq 1$ for all $x \in M$ and $\phi(q) = 1$ if $q \in D$ and $\phi(x) = 0$ if $x \in M - U$. Then we define

$$\tilde{X} = \begin{cases} \phi Y & \text{on } U, \\ 0 & \text{on } M - U. \end{cases}$$

Thus $\tilde{X}(p) = \phi(p)Y(p) = X$ and by construction $\tilde{X}$ is a C^∞-vector field on M.

Example (2) For the manifold $G = GL(V)$ we identified in Section 2.5, $T_I(G)$ with $g = gl(V)$ and for $X \in g$ we define a C^∞-vector field $\tilde{X}$ on G by its action on $f \in C^\infty(G)$ at $p \in G$

$$(\tilde{X}f)(p) = X(f \circ L(p)) = [TL(p)(I) \cdot X](f)$$

where $L(p) : G \to G : q \to pq$. Then $\tilde{X}$ is C^∞ since the right side of the equality consists of C^∞-operations and note $(\tilde{X}f)(I) = X(f)$ so that $\tilde{X}(I) = X$. Also

$\tilde{X}(p) = [TL(p)(I)]X \in T(G, p)$ so that $[\pi \circ \tilde{X}](p) = p$ which shows $\tilde{X}$ is a vector field. Using (*) of example (3), Section 2.5 we have

$$\begin{aligned}(\tilde{X}f)(p) &= X(f \circ L(p))\\ &= \lim_{t \to 0} \frac{[f \circ L(p)](\exp tX) - [f \circ L(p)](I)}{t}\\ &= \lim_{t \to 0} \frac{f(p \exp tX) - f(p)}{t}\\ &= \frac{d}{dt} f(p \exp tX)|_{t=0}.\end{aligned}$$

We shall let $D(M)$ denote the set of all C^∞-vector fields on M. Then we have the following algebraic results concerning these derivations

Proposition 2.30 (a) $D(M)$ is a Lie algebra over R relative to the **bracket multiplication** $[X, Y] = XY - YX$.

(b) $D(M)$ is a left F-module over the ring $F = C^\infty(M)$.

Proof (a) Clearly if $X, Y \in D(M)$ and $a, b \in R$, then $aX + bY \in D(M)$ by just checking the properties of a derivation. Next we shall show $[X, Y] = XY - YX$ is a derivation

$$\begin{aligned}[X, Y](fg) &= X[(Yf)g + f(Yg)] - Y[(Xf)g + f(Xg)]\\ &= (XYf)g + (Yf)(Xg) + (Xf)(Yg) + f(XYg)\\ &\quad - (YXf)g - (Xf)(Yg) - (Yf)(Xg) - f(YXg)\\ &= ([X, Y]f)g + f([X, Y]g).\end{aligned}$$

The multiplication $[X, Y]$ is bilinear and satisfies $[X, Y] = -[Y, X]$. Also the Jacobi identity

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

is a straightforward computation which is always satisfied for the bracket of endomorphisms.

(b) It is easy to see that the various defining properties of a left module are satisfied; for example, $(f + g)X = fX + gX$ or $(fg)X = f(gX)$ for $f, g \in F$ and $X \in D(M)$. However, note that $D(M)$ is not a "Lie algebra" over F since for "scalars" $f, g \in F$ we do not obtain the correct action relative to the product

$$[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X \neq fg[X, Y].$$

REMARK If for a chart (U, x) on M we let $X = \sum a_i \partial_i$ and $Y = \sum b_i \partial_i$ be in $D(U)$, then for any $f \in C^\infty(U)$, $(XY)(f) = X(Yf) \in C^\infty(U)$ and

$$\begin{aligned}(XY)(f) &= \sum_k X(b_k \partial_k(f)) \\ &= \sum_{k,i} \left(a_i \frac{\partial b_k}{\partial x_i} \frac{\partial}{\partial x_k} + b_k a_i \frac{\partial^2}{\partial x_i \partial x_k}\right)(f)\end{aligned}$$

which shows XY is not a tangent vector. However because the order of differentiation can be interchanged, the second-order derivatives vanish in $[X, Y]$.

Example (3) Let u_1, u_2, u_3 be coordinate functions on $M = R^3$ and let

$$X = u_2 \partial_3 - u_3 \partial_2, \qquad Y = u_3 \partial_1 - u_1 \partial_3, \qquad Z = u_1 \partial_2 - u_2 \partial_1.$$

Then X, Y, Z are linearly independent (over R) C^∞-vector fields and the vector space L spanned by X, Y, Z is a Lie algebra because the products

$$[X, Y] = -Z, \qquad [Y, Z] = -X, \qquad [Z, X] = -Y$$

are all in L.

Next we consider the action of a C^∞-map $f: M \to N$ on vector fields. First we note that f induces a map $Tf(p) : T(M, p) \to T(N, f(p))$ which maps tangent vectors into tangent vectors. However, in general, it is not possible to map vector fields on M into vector fields on N by Tf. Thus for any $X \in D(M)$ define the map

$$Tf(X) : M \to T(N) : p \to [Tf(p)]X(p)$$

noting that $[Tf(p)]X(p) \in T(N, f(p))$.

One would like to use $(Tf)X$ to define a vector field over N or even over $f(M)$ by taking a point $r = f(p) \in N$ and defining $[(Tf)X](r)$ to be $[Tf(p)]X(p)$. However, this is not always possible as shown by the following. Suppose $p \neq q$ but $f(p) = f(q)$. Let $X \in D(M)$ be a C^∞-vector field such that $Tf(p)X(p) \neq Tf(q)X(q)$ both of which are in $T(N, f(p))$. Then we can not assign a unique value to $(Tf)X$ at $r \in N$ by the desired process.

Exercise (1) If $f: M \to N$ is a diffeomorphism, then show that a vector field can be defined on N by the formula $[(Tf)X] \circ f^{-1} : N \to T(N)$.

Definition 2.31 Let $f: M \to N$ be a C^∞-map and let $X \in D(M)$, $Y \in D(N)$ be vector fields. Then X and Y are ***f*-related** if $(Tf)X = Y \circ f$; that is, for all $p \in M$, $Tf(p) \cdot X(p) = Y(f(p))$.

Thus if X and Y are f-related, then for every $g \in C^\infty(N)$

$$(Yg) \circ f = X(g \circ f),$$

for if $p \in M$, then

$$\begin{aligned}[(Yg) \circ f](p) = (Yg)(f(p)) = {} & Y(f(p))g \\ & = [Tf(p) \cdot X(p)]g = X(p)(g \circ f) = [X(g \circ f)](p).\end{aligned}$$

Definition 2.32 Let $f: M \to M$ be a C^∞-map and let $X \in D(M)$ be a vector field on M. Then X is **f-invariant** if X is f-related to X. Thus X "commutes" with the action of f by means of the formula $TfX = X \circ f$; that is $Tf(p)X(p) = X(f(p))$.

Another way of viewing the f-invariance of X is by noting that $Tf(p)X(p)$ and $X(f(p))$ are both in $T(M, f(p))$ so that the f-invariance of X means they are equal.

Example (4) For $G = GL(V)$ and $X \in gl(V)$ we defined the vector field $\tilde{X}$ on G by $(\tilde{X}g)(p) = X(g \circ L(p))$ where $g \in C^\infty(G)$. Now for any $a \in G$, $\tilde{X}$ is $L(a)$-invariant. Thus let $f = L(a)$, then for any $g \in C^\infty(G)$

$$\begin{aligned}[Tf(p)\tilde{X}(p)](g) &= \tilde{X}(p)(g \circ f) \\ &= [\tilde{X}(g \circ f)](p) \\ &= X((g \circ f) \circ L(p)), \qquad \text{definition of } \tilde{X} \\ &= X(g \circ L(a) \circ L(p)), \qquad \text{using } f = L(a) \\ &= X(g \circ L(ap)) = (\tilde{X}g)(ap) = [\tilde{X}(ap)](g).\end{aligned}$$

Thus $Tf(p)\tilde{X}(p) = \tilde{X}(f(p))$; that is, $TL(a)\tilde{X} = \tilde{X} \circ L(a)$. The vector field $\tilde{X}$ is called **left invariant** or **G-invariant** and will be used in yet another definition of the Lie algebra of G.

Proposition 2.33 Let X_1 and Y_1, X_2 and Y_2 be f-related. Then $[X_1, X_2]$ is f-related to $[Y_1, Y_2]$.

PROOF Since the X's and Y's are f-related we have using the paragraph following Definition 2.31 for any $g \in C^\infty(N)$ that

$$\begin{aligned}Y_2(Y_1 g)f(p) = {} & X_2(Y_1 g \circ f)(p) \\ & = X_2(X_1(g \circ f))(p) = [X_2 X_1(g \circ f)](p).\end{aligned}$$

Thus since a similar formula holds for $Y_1 Y_2$ we have

$$([Y_1, Y_2]g)f(p) = ([X_1, X_2](g \circ f))(p)$$

so that $[Y_1, Y_2]$ is f-related to $[X_1, X_2]$.

Proposition 2.34 Let $f: M \to N$ be a C^∞-map.

(a) If f is an immersion, then for every $Y \in D(N)$ there is at most one $X \in D(M)$ such that X and Y are f-related. In this case the $X \in D(M)$ exists if and only if for every $p \in M$ we have $Y(f(p)) \in Tf(p)T(M, p)$.

(b) If f is a surjection, then for every $X \in D(M)$ there is at most one $Y \in D(N)$ such that X and Y are f-related.

PROOF (a) Let $Y \in D(N)$ and let $X, Z \in D(M)$ with $Tf(p)X(p) = Y(f(p)) = Tf(p)Z(p)$. Then since $Tf(p)$ is injective $X(p) = Z(p)$; that is, $X = Z$. Now if X exists, then by definition $Y(f(p)) = Tf(p)X(p) \in Tf(p)T(M, p)$ and conversely one can define X by $X(p) = Tf(p)^{-1}Y(f(p))$ and this defines a vector field on M.

All that remains to show is that X is C^∞. Now since $Tf(p)$ is injective we have from Section 2.5 that for $p \in M$ there is a chart (V, y) at $f(p)$ in N so that (U, x) is a chart at p where $x_i = y_i \circ f$ for $i = 1, \ldots, m$. Now with these coordinates we let $X = \sum a_i \partial_i$ on U, then for $q \in U$ we have

$$\begin{aligned} a_i(q) = X(x_i)(q) &= [X(y_i \circ f)](q) \\ &= [(TfX)(y_i)](q) = [(TfX)(q)](y_i) \\ &= [Y(f(q))](y_i) = [(Yy_i) \circ f](q) \end{aligned}$$

which shows a_i is C^∞ because Yy_i and f are C^∞.

The proof of (b) is a straightforward exercise.

Exercises (2) Let $f: M \to M$ be a C^∞-map. Show that the set of f-invariant vector fields in $D(M)$ is a Lie subalgebra of $D(M)$.

(3) Let $G = GL(V)$ and let μ be an analytic multiplication on G; that is,

$$\mu : G \times G \to G : (x, y) \to \mu(x, y)$$

is an analytic mapping of manifolds. Now form the differential

$$T\mu : T(G, x) \times T(G, y) \to T(G, \mu(x, y));$$

that is, for $X \in T(G, x)$ and $Y \in T(G, y)$ we have

$$[(T\mu)(x, y)](X, Y) \in T(G, \mu(x, y)).$$

(i) For $X \in T_I(G) = gl(V)$ show the map

$$l(\mu, X) : G \to T(G) : x \to [(T\mu)(x, I)](0, X)$$

is an analytic vector field on G if and only if $\mu(x, I) = x$ for all $x \in G$;

(ii) Similarly discuss the function

$$r(\mu, X) : G \to T(G) : x \to [(T\mu)(I, x)](X, 0);$$

(iii) In case μ is the usual associative multiplication on G, compare the vector fields $l(\mu, X)$, $r(\mu, X)$, and $\tilde{X}$ of example (2).

(4) What can be said about a function $f \in C^\infty(R^m)$ such that $[fX, Y] = f[X, Y]$ for all C^∞-vector fields X, $Y \in D(R^m)$ (note Proposition 2.30)?

8. Integral Curves

Let α be a C^∞-curve defined on (a, b) into M as discussed in Section 2.5. Then the tangent vector $\dot{\alpha}(t)$ is given by $\dot{\alpha}(t) = [T\alpha(d/du)](t) \in T(M, \alpha(t))$. Thus $\dot{\alpha} : (a, b) \to T(M)$ is a C^∞-map such that the accompanying diagram is commutative.

$$\begin{tikzcd}
(a, b) \arrow[r, "\dot{\alpha}"] \arrow[dr, "\alpha"'] & T(M) \arrow[d, "\pi"] \\
 & M
\end{tikzcd}$$

Definition 2.35 Let M be an m-dimensional C^∞-manifold and let X be a C^∞-vector field on M. An **integral curve of** X is a C^∞-curve $\alpha : (a, b) \to M$ such that the tangent vector to α at each $t \in (a, b)$ equals the value of X at $\alpha(t)$; that is, $\dot{\alpha}(t) = X(\alpha(t))$ all $t \in (a, b)$. Thus the accompanying diagram is commutative.

$$\begin{tikzcd}
(a, b) \arrow[r, "\dot{\alpha}"] \arrow[dr, "\alpha"'] & T(M) \\
 & M \arrow[u, "X"']
\end{tikzcd}$$

In terms of a chart (U, x) of M we have from Section 2.5,

$$\dot{\alpha} = \sum d(x_i \circ \alpha)/dt\, \partial_i$$

and writing X in coordinates on U

$$X = \sum a_i\, \partial_i$$

we obtain α as an integral curve of X if and only if

$$d(x_i \circ \alpha)/dt = a_i(x_1 \circ \alpha, \ldots, x_m \circ \alpha) \qquad \text{for} \quad i = 1, \ldots, m.$$

We now summarize the facts we shall need concerning the solutions of such differential equations and the proofs can be found in the work of Dieudonné [1960] and Lang [1968].

Proposition 2.36 Let U be an open subset of R^m, let $p \in U$, and let $a_i \in C^\infty(U)$ for $i = 1, \ldots, m$. Then

(a) there exists an open neighborhood D of p with $D \subset U$;
(b) there exists an open interval $(-\varepsilon, \varepsilon) \subset R$;
(c) there exists a C^∞-map $f: (-\varepsilon, \varepsilon) \times D \to U : (t, w) \to f(t, w)$ such that for each $w \in D$ the function $\alpha_w : (-\varepsilon, \varepsilon) \to U : t \to f(t, w)$ with $\alpha_i = u_i \circ \alpha_w$ for $i = 1, \ldots, m$ satisfy

(i) $d\alpha_i/du(t) = a_i(\alpha_1(t), \ldots, \alpha_m(t))$ all $t \in (-\varepsilon, \varepsilon)$, and
(ii) $\alpha_i(0) = w_i$ where $w_i = u_i(w)$.

Moreover if $\bar{\alpha}_w : (-\bar{\varepsilon}, \bar{\varepsilon}) \to U$ with $(-\bar{\varepsilon}, \bar{\varepsilon}) \subset (-\varepsilon, \varepsilon)$ satisfies (i) and (ii), then $\bar{\alpha}_w = \alpha_w | (-\bar{\varepsilon}, \bar{\varepsilon})$.

Thus the unique solutions to the above differential equations depend in a C^∞-manner on the initial conditions. We now translate these facts to manifolds [Bishop and Goldberg, 1968; Singer and Thorpe, 1967].

Theorem 2.37 (a) Let M be a C^∞-manifold and $X \in D(M)$ a C^∞-vector field on M and let $p \in M$. Then there exists an open neighborhood D of p in M and an open interval $(-\varepsilon, \varepsilon) \subset R$ and a C^∞-map f: $(-\varepsilon, \varepsilon) \times D \to M$ such that for each $w \in D$ the curve

$$\alpha_w : (-\varepsilon, \varepsilon) \to M : t \to f(t, w)$$

is the unique local integral curve of X defined on $(-\varepsilon, \varepsilon)$ with $\alpha_w(0) = w$. In particular α_p is a local integral curve through $p \in M$.

(b) For each $t \in (-\varepsilon, \varepsilon)$ the C^∞-map $\phi(t)$ given by $\phi(t) : D \to M : w \to f(t, w)$ satisfies:

(i) if s, t and $s + t$ are in $(-\varepsilon, \varepsilon)$, then $\phi(s + t) = \phi(s) \circ \phi(t)$ on $\phi(t)^{-1}(D) \cap D$;
(ii) if $t \in (-\varepsilon, \varepsilon)$, then $\phi(t)^{-1}$ exists on $D \cap \phi(t)(D)$ and $\phi(t)^{-1} = \phi(-t)$.

A map $\phi : (-\varepsilon, \varepsilon) \times D \to M$ such that $\phi(t)$ satisfies (i) and (ii) above is called a **local one-parameter group on** M.

PROOF (a) Let (U, x) be a chart at p in M and let $U' = x(U) \subset R^m$. Then on U' we have $X = \sum a_i \partial_i$ where $a_i \in C^\infty(U')$. By Proposition 2.36, there exist $D' \subset U'$ and $(-\varepsilon, \varepsilon) \subset R$ and $f' : (-\varepsilon, \varepsilon) \times D' \to U'$ with the desired properties which can be translated back to M by x^{-1}.

(b) We use the uniqueness part of Proposition 2.36 as follows. For fixed $t \in (-\varepsilon, \varepsilon)$ the curves $u(s) = f(s + t, w)$ and $v(s) = f(s, \phi(t)(w))$ are integral curves of X defined on a subinterval of $(-\varepsilon, \varepsilon)$ which contains 0 and by the initial conditions we have $u(0) = v(0) = f(t, w)$. Thus by the uniqueness $u = v$; that is, $\phi(s + t) = \phi(s) \circ \phi(t)$. Also $\phi(t)^{-1} = \phi(-t)$.

The preceding results on differential equations are also true when C^∞ is replaced by "analytic." Furthermore, if the vector field depends analytically upon a parameter, then the integral curve does also as follows.

Definition 2.38 Let M be an analytic manifold and let V be a Euclidean vector space over R. Let A denote an element in V and let $X(A)$ be an analytic vector field which is a function of $A \in V$. Then $X(A)$ **depends analytically on the parameter** $A \in V$ if for any $p \in M$ and any function f analytic at p, the mapping $\mathrm{Dom}(f) \times V \to R : (q, A) \to [X(A)(f)](q)$ is analytic.

Using the results of Dieudonné [1960] and Lang [1968] on this dependence we have the following.

Theorem 2.39 Let M be an analytic manifold, V a Euclidean vector space over R, and $X(A)$ an analytic vector field which depends analytically upon the parameter $A \in V$. Then for any $p \in M$, there exist an open interval $(-\varepsilon, \varepsilon) \subset R$ and an open convex neighborhood U of 0 in V and an analytic map $u : (-\varepsilon, \varepsilon) \times U \to M : (t, A) \to u(t, A)$ which is the unique local integral curve of $X(A)$ through $p \in M$.

PROOF Since this is a local result, we can assume M is an open set in R^m so that the vector field $X(A)$ can be represented by analytic functions $a_i(x, A)$ on $M \times V$ for $i = 1, \ldots, m$; that is, $X(A) = \sum a_i \partial$ where $a_i : M \times V \to R$ are analytic. Thus we now have as before a system of (parameterized) differential equations for the integral curve, and the results follow from Dieudonné [1960, Theorem 10.7.5], for example.

Exercise (1) Show that the vector fields $\tilde{X}, l(\mu, X)$, and $r(\mu, X)$ in exercise (3), Section 2.7 depend analytically on the parameter $X \in gl(V)$.

Example (1) Let $p = (0, 0) \in R^2 = M$ with coordinates (u_1, u_2) and let $X = \partial_1 + \exp(-u_2)\,\partial_2$ be a vector field on all of M. Then the equation for the integral curve α is

$$d\alpha_1/dt = 1 \qquad \text{and} \qquad d\alpha_2/dt = \exp(-\alpha_2).$$

Let $D = \{(x, y) \in R^2 : -1 < y < 1\}$ be an open neighborhood of p and let $\varepsilon = e^{-1}$. Then for $t \in (-\varepsilon, \varepsilon)$ and for $w = (w_1, w_2) \in D$, the C^∞-map

$$f : (-\varepsilon, \varepsilon) \times D \to M : (t, w) \to \big(t + w_1, \log(t + \exp w_2)\big)$$

is such that

$$\alpha_w(t) = \big(\alpha_1(t), \alpha_2(t)\big)$$

with

$$\alpha_1(t) = t + w_1 \qquad \text{and} \qquad \alpha_2(t) = \log(t + \exp w_2)$$

is a solution to the above equation with $\alpha_w(0) = w$.

REMARK Theorem 2.37 gives only local existence and uniqueness of integral curves and it is not always possible to find global curves; that is, it is not always possible to extend the domain $(-\varepsilon, \varepsilon)$ to all of R. Thus, for example, let $M = R^2 - \{(0, 0)\}$ with coordinates (u_1, u_2) and let $X = \partial_1$. Then the integral curve $\alpha(t)$ of X through $(1, 0)$ is $\alpha(t) = (t + 1, 0)$ which cannot be extended to a curve in M defined on all of R because $(0, 0)$ is not in M.

Let $M = R^2$ and let $X = -u_2\,\partial_1 + u_1\,\partial_2$ be a vector field on M. Then the general form for the integral curve $\alpha_w(t)$ is

$$\alpha_w(t) = (w_1 \cos t - w_2 \sin t, w_2 \cos t + w_1 \sin t)$$

and $\alpha_w(0) = w$. Note that $\alpha_w(t)$ is defined for all $t \in R$.

Definition 2.40 A vector field is **complete** if all its integral curves have domains all of R.

Exercise (2) Show $X = -u_2\,\partial_1 + u_1\,\partial_2$ is complete on R^2. Is $X = \exp(-u_1)\,\partial_1 + \partial_2$ complete on R^2?

Examples (2) Let $G = GL(V)$ and let $g = gl(V)$ be identified with $T_I(G)$. For $X \in g$ we have defined the G-invariant vector field $\tilde{X}$ by $(\tilde{X}f)(a) = X(f \circ L(a))$ for all $a \in G$ and $f \in C^\infty(G)$. Let E_{ij} be the usual matrix basis of $\mathrm{End}(V)$ which gives coordinate functions u_{ij} on G; that is, $u_{ij}(a) = (a_{ij})$. We write $\tilde{X} = \sum X_{ij}\,\partial/\partial u_{ij}$ so that $X_{ij} = \tilde{X}(u_{ij})$ are in $C^\infty(G)$ and we now compute the coordinate functions X_{ij}.

For $a, x \in G$ we have

$$(u_{ij} \circ L(a))(x) = u_{ij}(ax) = \sum_{k=1}^{m} u_{ik}(a)u_{kj}(x),$$

using matrix multiplication. Thus applying X to this formula for $u_{ij} \circ L(a)$ we have

$$\begin{aligned} X(u_{ij} \circ L(a)) &= X(\sum_k u_{ik}(a)u_{kj}) \\ &= \sum_k u_{ik}(a)X(u_{kj}) = \sum_k u_{ik}(a)x_{kj}, \end{aligned}$$

where we write $X = \sum x_{ij}\, \partial/\partial u_{ij}(I) \in g$ and have $x_{pq} = X(u_{pq}) = X_{pq}(I)$. Thus letting $f = u_{ij}$ in the definition of $\tilde{X}$ we obtain

$$\begin{aligned} X_{ij}(a) &= (\tilde{X}u_{ij})(a) \\ &= X(u_{ij} \circ L(a)) = \sum_k u_{ik}(a)x_{kj}, \end{aligned} \tag{$*$}$$

so that the equation for an integral curve α of $\tilde{X}$ is

$$d(u_{ij} \circ \alpha)/ds = \sum_k (u_{ik} \circ \alpha)x_{kj} \qquad \text{for} \quad i = 1, \ldots, m.$$

From example (2) in Section 2.7 on $\tilde{X}$ we have

$$(\tilde{X}f)(p) = df(p \exp tX)/dt\,|_{t=0}$$

so that for $p = q \exp sX$ we have

$$\begin{aligned} (\tilde{X}f)(q \exp sX) &= df[q \cdot \exp(s + t)X]/dt\,|_{t=0} \\ &= df(q \cdot \exp uX)/du\,|_{u=s} = df(q \cdot \exp sX)/ds, \end{aligned}$$

where associativity is used in the first equality. Thus for $q = I$ and $f = u_{ij}$ we have

$$du_{ij}(\exp sX)/ds = (\tilde{X}u_{ij})(\exp sX) = \sum_k u_{ik}(\exp sX)x_{kj}$$

using ($*$) above for the last equality. This shows that $\alpha(s) = \exp sX$ is the solution of the equation for the integral curve of $\tilde{X}$

$$d\alpha_{ij}/ds = \sum_k \alpha_{ik}\, x_{kj} \qquad \text{and} \qquad \alpha(0) = I,$$

where $\alpha_{ij} = u_{ij} \circ \alpha$. In terms of the given matrix $X = (X_{ij})$ this equation can be written: $d\alpha/ds = \alpha X$ which yields $\alpha(s) = \exp sX$ which is a one-parameter group defined on all of R. If the initial condition is changed to $\alpha(0) = A$, then for $\tilde{X}$ we obtain the integral curve $\alpha_A(s) = A \cdot \exp sX$ and $\alpha_A(0) = A$. From this we see $\tilde{X}$ is a complete vector field on G.

(3) We now consider a Taylor's series expansion for a real-valued analytic function f on the analytic manifold $G = GL(V)$. Thus let $X \in g$ and $\tilde{X}$ be as in the preceding example and let f be analytic at $p \in G$. Then from this example we have

$$df(p \exp sX)/ds = (\tilde{X}f)(p \exp sX) = [\tilde{X}(p \exp sX)](f)$$

and by induction

$$d^nf(p \exp sX)/ds^n = [\tilde{X}(p \exp sX)](\tilde{X}^{n-1}(f)) = (\tilde{X}^nf)(p \exp sX).$$

Thus if we write $g(s) = f(p \exp sX)$, we have, since g is the composition of analytic functions, the power series in a suitable interval containing $0 \in R$

$$g(s) = \sum \frac{a_n}{n!} s^n,$$

where the $a_n \in R$ are computed by differentiation as usual

$$a_n = d^n f(p \exp sX)/ds^n |_{s=0} = (\tilde{X}^n f)(p).$$

Thus if we define the operator formula

$$[(\exp s\tilde{X})(f)](p) = \sum_{n=0}^{\infty} \frac{s^n \tilde{X}^n f}{n!}(p),$$

we obtain the following version of Taylor's formula for $GL(V)$

$$f(p \exp sX) = [(\exp s\tilde{X})(f)](p).$$

Exercise (3) (i) Consider the C^∞-vector field on R^3 defined by

$$X(p) = p_2\,(\partial/\partial x_1)(p) + p_3\,(\partial/\partial x_2)(p) + p_1\,(\partial/\partial x_3)(p)$$

where $p = (p_1, p_2, p_3)$. Find the integral curve $\alpha(t)$ of X so that $\alpha(0) = (-1, 1, 1)$.

(ii) Let the C^∞-vector field on R^3 be given by $Y(p) = p_1 p_2\,(\partial/\partial x_3)(p)$. Compute $[X, Y](p)$.

CHAPTER 3

TOPOLOGICAL GROUPS

In our previous discussion of some matrix groups it was observed that we were studying not only the group operations but also the continuity of these operations. Thus in this chapter we abstract the situation and consider groups which are topological spaces so that the group operations are continuous relative to the topology of the space. We then prove facts for these topological groups which indicate that much information can be obtained from a neighborhood of the identity element; this leads to local groups and local isomorphisms. Next we consider topological subgroups, coset spaces, and normal subgroups. Finally, for connected topological groups, we show that any neighborhood of the identity actually generates the group as an abstract group.

1. Basics

In the next chapter, we shall apply the results of the preceding chapters to obtain elementary results on Lie groups. However, since a Lie group is a topological group, we shall briefly discuss this more general situation.

Definition 3.1 A **topological group** is a set G such that:

(a) G is a Hausdorff topological space;
(b) G is a group;
(c) the mappings $G \times G \to G : (x, y) \to xy$ and $G \to G : x \to x^{-1}$ are continuous, where $G \times G$ has the product topology.

Thus the set G has two structures—topological and algebraic—and they are related by property (c); that is, the group structure is **compatible** with the topological structure.

The compatibility conditions in (c) are equivalent to the following single condition:

(c′) the mapping $G \times G \to G : (x, y) \to xy^{-1}$ is continuous.

This condition holds for if (c) holds, then we have that $G \times G \to G \times G : (x, y) \to (x, y^{-1})$ is continuous. Consequently the map $G \times G \to G : (x, y) \to (x, y^{-1}) \to xy^{-1}$ is continuous. Conversely if (c′) holds, then set $x = e$ (the identity) to obtain $y \to (e, y) \to ey^{-1} = y^{-1}$ is a continuous map. Also from $xy = x(y^{-1})^{-1}$the map $(x, y) \to xy$ is continuous.

We can express (c) in terms of neighborhoods as follows. For any $x, y \in G$ and for any neighborhood W of xy in G, there exist neighborhoods U of x and V of y with $UV \subset W$. Also for any neighborhood U of x^{-1}, we have $U^{-1} = \{a^{-1} : a \in U\}$ is a neighborhood of x. Thus replacing x by x^{-1}, we have if V is a neighborhood of x, then V^{-1} is a neighborhood of x^{-1}.

Definition 3.2 Let G be a topological group and let $a \in G$. Then the map

$$L(a) : G \to G : x \to ax$$

is called a **left translation**. Similarly the map $R(a) : G \to G : x \to xa$ is called a **right translation.**

It should be noted that the maps $L(a)$ and $R(a)$ for $a \in G$ are homeomorphisms of G. Furthermore given any two points $x, y \in G$, then the homeomorphism $L(yx^{-1})$ maps x onto y. In particular, there always exists a homeomorphism which maps $e \in G$ onto any other element $a \in G$ and using this, we shall see many of the local properties of $a \in G$ are determined by those of e. Thus, for example, U is a neighborhood of $a \in G$ if and only if $U = L(a)V = aV$ where V is a neighborhood of $e \in G$.

Proposition 3.3 Let G be a topological space which is also a group. Then G is a topological group relative to these two structures if and only if:

(a) the set $\{e\}$ is closed;
(b) for all $a \in G$ the translations $R(a)$ and $L(a)$ are continuous;
(c) the mapping $G \times G \to G : (x, y) \to xy^{-1}$ is continuous at the point (e, e).

PROOF Let $a, b \in G$ with $a \neq b$. Then we shall find disjoint neighborhoods of a and b as follows. Since $L(a)^{-1} = L(a^{-1})$ is continuous, $L(a)$ is a homeomorphism. Thus $\{a\} = L(a)\{e\}$ and $\{b\} = L(b)\{e\}$ are closed, and there exists a neighborhood U of b such that $U \cap \{a\}$ is empty. Otherwise if $U \cap \{a\}$ is not empty for every neighborhood U of b, then since $\{a\}$ is closed, $b = a$. Now using (b) the neighborhood $U = bV$ where V is a neighborhood of e. Thus from (c) and using $e = ee^{-1}$ we can find a neighborhood W of e with $WW^{-1} \subset V$. We shall now show that $aW \cap bW$ is empty, and therefore G is Hausdorff. Thus suppose $x \in aW \cap bW$. Then $x = aw_1 = bw_2$ with $w_i \in W$ and therefore

$$a = bw_2w_1^{-1} \in (bW)W^{-1} \subset bV \subset U$$

which contradicts $U \cap \{a\}$ being empty.

Next we shall show that the topology and group operations are compatible by showing the map $G \times G \to G : (x, y) \to xy^{-1}$ is continuous at any point $(a, b) \in G \times G$. First we note that $R(a)^{-1} = R(a^{-1})$ is continuous, and therefore $R(a)$ is a homeomorphism. Now let W be a neighborhood of ab^{-1}. Then $a^{-1}Wb = R(b)L(a)^{-1}W$ is a neighborhood of e. Using the fact that $(x, y) \to xy^{-1}$ is continuous at (e, e), we let U and V be neighborhoods of e so that $UV^{-1} \subset a^{-1}Wb$. Then we have for the neighborhoods aU and bV of a and b that

$$(aU)(bV)^{-1} = aUV^{-1}b^{-1} \subset a(a^{-1}Wb)b^{-1} = W,$$

and this shows continuity. The converse follows from various preceding remarks and is left as an exercise.

Lemma 3.4 Let U be a neighborhood of e in a topological group G. Then there exists a neighborhood V of e such that $V \subset U$, $V = V^{-1}$ $(= \{v^{-1} : v \in V\})$ and $VV = VV^{-1} \subset U$. We shall call such a neighborhood V of e **symmetric**. Furthermore, in this case $\overline{V} \subset U$ where $\overline{V}$ is the closure of V.

PROOF Since multiplication $G \times G \to G$ is continuous, there exists neighborhoods P and Q of e in G such that $PQ \subset U$. Now let $W = P \cap Q$ and $V = W \cap W^{-1}$. Then V is a neighborhood of e with $V = V^{-1}$ and also $VV \subset PQ \subset U$. Next let $x \in \overline{V}$. Then xV is a neighborhood of x and consequently $V \cap xV$ is not empty. Thus for some v, $v_1 \in V$ we have $xv = v_1$ and therefore $x = v_1v^{-1} \in VV^{-1} \subset U$.

Exercise (1) Show in detail that the neighborhoods U and V in the above proof are such that $V \subset U$.

Definition 3.5 A subset of a topological group G which contains an (open) neighborhood of the identity e is called a **nucleus** of G.

Now since the topology of a topological group G is determined by the family of neighborhoods at each of its points, we see by using the left or right translations that the topology is determined by the family of nuclei of G as follows [Chevalley, 1946; Cohn, 1957]:

Proposition 3.6 Let $\mathscr{V}$ be the family of all nuclei of a topological group G. Then $\mathscr{V}$ satisfies:

(a) $V_1, V_2 \in \mathscr{V}$ implies $V_1 \cap V_2 \in \mathscr{V}$;
(b) $V_1 \in \mathscr{V}$ and $V_1 \subset W \subset G$ implies $W \in \mathscr{V}$;
(c) for any $V_1 \in \mathscr{V}$, there exists $V \in \mathscr{V}$ such that $VV^{-1} \subset V_1$;
(d) if $V \in \mathscr{V}$ and $a \in G$, then $aVa^{-1} \in \mathscr{V}$;
(e) $\bigcap \{V : V \in \mathscr{V}\} = \{e\}$.

Conversely, given a group G and a family of subsets $\mathscr{V}$ of G satisfying (a)–(e), then there exists a unique topology for G relative to which G becomes a topological group and $\mathscr{V}$ is exactly the family of nuclei for this topological group.

PROOF Properties (a)–(e) are immediate and the converse can be regarded as a straightforward exercise. For example, define the topology on G by saying that W is open if $x \in W$ implies there exists $V \in \mathscr{V}$ and $xV \subset W$. The family of such sets W satisfies the axioms of a topology which makes $G \times G \to G : (x, y) \to xy^{-1}$ continuous. This topology is Hausdorff since $\{e\}$ is closed: If $a \neq e$, then $a^{-1} \neq e$ and by (e) there exists $V \in \mathscr{V}$ with $a^{-1} = a^{-1} \cdot e \notin V$. Thus $e \notin aV$ so that a is not in the closure of $\{e\}$; that is, $\{e\}$ is closed.

Examples (1) The matrix groups of the preceding chapters are topological groups, as are **discrete groups**: Let G be any group and let the topology be such that every subset of G is open; that is, the discrete topology. Then G is a Hausdorff space, and since $G \times G$ has the product topology (which is discrete), any map $G \times G \to G$ is continuous.

(2) The additive group of R with the usual metric topology is a topological group. However, if the topology is changed to another topology where the half-open intervals $[a, a + \varepsilon)$ with $\varepsilon > 0$ are taken to be a neighborhood basis at $a \in R$, then the operation $R \to R : x \to -x$ is not continuous at $0 \in R$ so R, with this topology, is not a topological group.

(3) Let G_1 and G_2 be topological groups. Then the product space $G_1 \times G_2$ with the product topology and the pointwise operations $(x_1, x_2)(y_1, y_2)^{-1} = (x_1 y_1{}^{-1}, x_2 y_2{}^{-1})$ becomes a topological group called the **product group** of G_1 and G_2. This example can be generalized to the **semidirect product** of G_1 and G_2 as follows [Hochschild, 1965]. Let ϕ be a homomorphism of G_2 into the automorphism group of G_1 denoted by $\mathrm{Aut}(G_1)$. On the product

space $G_2 \times G_1$, put the product topology and require that the map $G_2 \times G_1 \to G_1 : (y, x) \to \phi(y)x$ be continuous. Then $G_1 \times G_2$ with the product topology and product defined by

$$(x, y)(x_1, y_1) = (x[\phi(y)x_1], yy_1)$$

is a topological group called the "semidirect product" and denoted by $G_1 \times_\phi G_2$. When the homomorphism $\phi : G_2 \to \mathrm{Aut}(G_1)$ is such that $\phi(y)$ is the identity, then we obtain the direct product.

As an example of a product group we note that the unit circle $S^1 = \{e^{2\pi i x} : x \in R\}$ is a topological group relative to multiplication in the complex numbers. Then the torus $T^n = S^1 \times \cdots \times S^1$ as given in Section 2.1 can be regarded as a topological product group.

Exercise (2) Let $G = G_1 \times_\phi G_2$ be the semidirect product as in the above example (3).

(a) Show the multiplication of G is associative.
(b) What is the inverse of (x, y) in G?
(c) Show $H_1 = G_1 \times_\phi \{e\} = \{(x, e) : x \in G_1\}$ is a normal subgroup of G.
(d) Is $H_2 = \{e\} \times_\phi G_2$ necessarily a normal subgroup of G?

Definition 3.7 A **local group** is a Hausdorff space N such that:

(a) there is a binary operation in N, $(x, y) \to xy$ which is defined for certain pairs $(x, y) \in N \times N$;

(b) the operation is associative when defined. Thus if $x, y, z \in N$ and $(xy)z, x(yz) \in N$, then $(xy)z = x(yz)$;

(c) there exists an identity element $e \in N$. Thus for all $x \in N$, xe and ex are defined and $xe = ex = x$;

(d) there exists an inverse operation in N, $x \to x^{-1}$ which is defined for certain elements $x \in N$ such that if x^{-1} is defined, then xx^{-1} and $x^{-1}x$ are defined and $xx^{-1} = x^{-1}x = e$;

(e) the maps $(x, y) \to xy$ and $x \to x^{-1}$ are continuous where defined.

Thus if $xy = z$ is defined in N, then for any neighborhood U of z in N there exist neighborhoods V of x and W of y in N such that $VW = \{vw : v \in V \text{ and } w \in W\}$ is defined and $VW \subset U$. Similarly for the map $x \to x^{-1}$.

REMARK (1) Any open nucleus of a topological group is a local group, and we use local groups later in discussing Lie groups.

Analogous to the proof of Lemma 3.4, we have the following result:

Proposition 3.8 Let N be a local group. Then there exists a neighborhood U of e in N such that:

(a) $(xy)z = x(yz)$ all $x, y, z \in U$;
(b) $xx^{-1} = x^{-1}x = e$ all $x \in U$;
(c) $ex = xe = x$ all $x \in U$;
(d) for all $x \in U$ we have $x^{-1} \in U$;

where all the above products and inverses actually exist in N. A neighborhood U satisfying these conditions is called a **germ** of the local group N.

Definitions 3.9 (a) The local groups N and N' are **topologically isomorphic** if there exists a homeomorphism $f: N \to N' : x \to f(x)$ such that the product xy is defined in N if and only if the product $f(x)f(y)$ is defined in N' and in this case $f(xy) = f(x)f(y)$.

(b) The topological groups G and G' are **locally isomorphic** if they have open nuclei which as local groups are topologically isomorphic.

Example (4) The topological groups R and the torus T^1 have neighborhoods N and N^1 of the respective identities which are topologically isomorphic as local groups. Thus R and T^1 are locally isomorphic but not isomorphic as groups.

REMARK (2) In the definition of a topological group, we assumed the topological space G to be Hausdorff. However, using the group structure of G, we can start with weaker separation axioms for G and obtain stronger separation theorems than being Hausdorff. Good accounts of these theorems can be found in the work of Hewitt and Ross [1963] and Montgomery and Zippin [1955], and we now summarize some of the results.

Definitions (a) A topological space M is a **T_0-space** if for any given pair of distinct points $x, y \in M$, there exists an open set U of M which contains one of these points but not the other. A **T_0-topological group** is a group G which is a T_0-space and such that the map $G \times G \to G : (x, y) \to xy^{-1}$ is continuous.

(b) A metric (or pseudo-metric) d on a group G is **left invariant** (respectively **right invariant**) if for all $a, x, y \in G$ we have $d(ax, ay) = d(x, y)$ [respectively $d(xa, ya) = d(x, y)$]. If d is both left and right invariant, then d is called **two-sided invariant** or just **invariant**.

Theorem Let G be a T_0-topological group. Then G is metrizable if and only if there is a countable (open) basis at the identity $e \in G$. If this is the case, then the metric can be taken to be left invariant.

Corollary Let G be a T_0-topological group such that every point of G has a neighborhood U so that $\bar{U}$ is countably compact; that is, every countable open covering admits a finite subcovering. Then G has a left invariant metric which yields the original topology if and only if $\{e\}$ equals the intersection of a countable family of open sets.

Corollary Let G be a T_0-topological group which is compact and such that $\{e\}$ equals the intersection of a countable family of open sets. Then G has an invariant metric which yields the original topology of G.

Definition A topological space M is a T_1-**space** if for distinct points $x \neq y$ in M, there exists an open set U with $x \in U$ but $y \notin U$, and there exists an open set V with $y \in V$ but $x \notin V$.

Exercise (3) Show that a T_0-topological group is a T_1-space.

Definitions Let M be a T_1-space. Then:

(a) M is **regular** if for every closed set F in M and every $x \notin F$, there exist disjoint open sets U and V such that $x \in U$ and $F \subset V$;

(b) M is **completely regular** if given any $x \in M$ and given any closed set F with $x \notin F$, there exists a continuous function $g : M \to [0, 1]$ such that $g(x) = 0$ and $g(F) = 1$;

(c) M is **normal** if for every pair of disjoint closed sets F_1 and F_2 in M, there exist disjoint open sets U_1 and U_2 of M such that $F_1 \subset U_1$ and $F_2 \subset U_2$;

(d) M is **paracompact** if every open covering of M has a locally finite refinement. Recall that an open covering $\mathscr{U}$ is locally finite if for each $x \in M$, there exists an open set $V(x)$ which contains x and such that $\{U \in \mathscr{U} : U \cap V(x) \neq \phi\}$ is a finite set. Also recall that a locally finite refinement means that if $\mathscr{U}$ is any open covering, then there exists a locally finite open covering $\mathscr{V}$ such that $V \in \mathscr{V}$ implies there exists $U \in \mathscr{U}$ with $V \subset U$.

We have (d) implies (c) implies (b) implies (a).

Theorem Let G be a T_0-topological group. Then G is completely regular and consequently Hausdorff.

Theorem Let G be a locally compact T_0-topological group. Then G is paracompact and consequently normal.

There exist T_0-topological groups which are not normal topological spaces [Husain, 1966].

2. Subgroups and Homogeneous Spaces

We shall consider subgroups H of a topological group G and the corresponding space of left cosets $G/H = \{aH : a \in G\}$. Then we eventually consider the case when H is a normal subgroup so that G/H becomes a topological group.

Definition 3.10 Let G be a topological group and let H be a subset of G such that $HH^{-1} \subset H$. Then H is a **subgroup** of G (in the abstract sense). The topology of G induces a topology on the subgroup H by requiring $U \subset H$ to be open if and only if $U = H \cap V$ where V is open in G. If with this induced topology, the subgroup H becomes a topological group we call H a **topological subgroup**.

REMARK (1) We shall be interested in closed subgroups H of G; that is, H is closed as a subset of G. However, we note that if H is an open subgroup of G, then H is closed. For, since H is open, so is aH for all $a \in G$. Therefore $K = \bigcup\{aH : a \notin H\}$ is open so that the complement of K, which is H, is closed.

Exercise (1) If H is a subgroup of the topological group G, then show its closure $\bar{H}$ is also a subgroup of G. More generally, one can show that for subsets A, B of a topological group G that $\bar{A}\,\bar{B} \subset \overline{AB}$, $(\bar{A})^{-1} = (\overline{A^{-1}})$ and $a\bar{A}b = \overline{aAb}$ for all $a, b \in G$.

Theorem 3.11 Let H be a topological subgroup of the topological group G and let $\pi : G \to G/H : a \to aH$ be the natural projection. Then

(a) G/H can be made into a topological space such that:

(i) the projection $\pi : G \to G/H$ is continuous, and

(ii) if N is a topological space and if $f : G/H \to N$ is such that $f \circ \pi : G \to N$ is continuous, then f is continuous.

The topology defined on G/H is uniquely determined by (i) and (ii) and is called the **quotient topology**.

(b) G/H with the quotient topology is such that π is an open map; that is, U is open in G implies $\pi(U)$ is open in G/H.

(c) The quotient topology is Hausdorff if and only if H is a closed subset of G.

PROOF We define the topology on G/H by requiring a subset U of G/H to be open in G/H if and only if the inverse image $\pi^{-1}(U)$ is open in G. Then the axioms for open sets of a topology in G/H are satisfied. Furthermore, if U

is open in G/H, then by definition $\pi^{-1}(U)$ is open in G so that π is continuous. To prove (ii), let $f \circ \pi : G \to N$ be continuous and let W be open in N. Then $(f \circ \pi)^{-1}(W) = \pi^{-1}(f^{-1}(W))$ is open in G. Thus by definition of the quotient topology, $f^{-1}(W)$ is open in G/H so that f is continuous.

Let $\mathscr{T}$ be the topology given above on G/H and let $\mathscr{T}'$ be any other topology on G/H satisfying (i) and (ii). Let f be the identity map of G/H with topology $\mathscr{T}$ onto G/H with topology $\mathscr{T}'$. Then $f \circ \pi$ is continuous since $f \circ \pi(a) = aH = \pi(a)$ and $\mathscr{T}'$ satisfies (i) for $\pi : G \to G/H$. Thus since G/H with topology $\mathscr{T}$ satisfies (ii) where $N = (G/H, \mathscr{T}')$, we see that f is continuous. Interchanging the roles of $\mathscr{T}$ and $\mathscr{T}'$, we obtain f^{-1} is continuous so that the identity map f is a homeomorphism; that is, $\mathscr{T} = \mathscr{T}'$.

Next, to show that π is an open map, we note that if U is open in G, then Ua is open for all $a \in G$. Consequently $UH = \bigcup\{Ua : a \in H\}$ is open in G. However, since $UH = \pi^{-1}[\pi(U)]$, we have by the definition of the quotient topology that $\pi(U)$ is open in G/H.

For (c), we note that G/H being Hausdorff yields the fact that $\{eH\}$ is closed in G/H. This implies $H = \pi^{-1}\{eH\}$ is closed in G because π is continuous. Conversely, suppose H is closed in G and let $aH \neq bH$ in G/H. Then $a \notin bH$. Thus since bH is closed, there exists a neighborhood U of e in G such that Ua is a neighborhood of a in G and $Ua \cap bH$ is empty. From Lemma 3.4, there is a neighborhood V of e such that $V^{-1}V \subset U$ and consequently $(Va)H = V(aH)$ and $V(bH)$ are neighborhoods of aH and bH, respectively. This uses $\pi : G \to G/H$ as an open map and Va and Vb as neighborhoods of a and b in G. These neighborhoods are disjoint, for if $p \in V(aH) \cap V(bH)$, then $p = vah = v_1bh_1$ for $v, v_1 \in V$ and $h, h_1 \in H$. Therefore $q = v_1^{-1}va = bh_1h^{-1}$ is in $Ua \cap bH$, a contradiction.

Definition 3.12 A subgroup H is a **normal subgroup** of G if $aHa^{-1} \subset H$ all $a \in G$ and then G/H is a group relative to $aH \cdot bH = abH$ which is called the **quotient group.**

Corollary 3.13 Let H be a closed normal subgroup of the topological group G and let G/H be the quotient group. Then relative to the quotient topology, G/H becomes a topological group such that the projection $\pi : G \to G/H$ is an open continuous homomorphism.

PROOF It suffices to show that

$$G/H \times G/H \to G/H : (aH, bH) \to ab^{-1}H$$

is continuous. Let U be a neighborhood of $ab^{-1}H = \pi(ab^{-1})$ in G/H. Then $\pi^{-1}(U)$ is a neighborhood of ab^{-1} in G. Now there exist neighborhoods V of a and W of b in G such that $VW^{-1} \subset \pi^{-1}(U)$. However, since π is open, $\pi(V)$

and $\pi(W)$ are neighborhoods of $aH = \pi(a)$ and $bH = \pi(b)$, respectively. Thus $\pi(V)[\pi(W)]^{-1} = \pi(VW^{-1}) \subset U$ which proves continuity.

Corollary 3.14 If H is an open normal subgroup of the topological group G, then G/H is discrete.

PROOF Since H is open, the cosets aH for $a \in G$ are open in G. Thus since π is an open map, the sets $\{aH\}$ in G/H are open. Therefore G/H is discrete.

Corollary 3.15 Let $f: G \to \bar{G}$ be a homomorphism of topological groups. Then f is continuous if and only if f is continuous at the identity $e \in G$.

PROOF Assume f is continuous at the identity. Let $a \in G$ and let $f(a)\bar{U}$ be a neighborhood of $f(a)$ in $\bar{G}$ where $\bar{U}$ is a neighborhood of $\bar{e}$ in $\bar{G}$. Since f is continuous at $e \in G$ and since $\bar{e} = f(e)$, there exists a neighborhood U of e in G such that $f(U) \subset \bar{U}$ which proves continuity at a since aU is a neighborhood of a in G with $f(aU) \subset f(a)\bar{U}$.

Using the preceding results with the isomorphism theorem for groups, we have the following result which we leave as an exercise [Cohn, 1957].

Theorem 3.16 Let $f: G \to \bar{G}$ be a continuous homomorphism of the topological groups G and $\bar{G}$ and let $H = \{x \in G : f(x) = \bar{e}\}$ be the kernel of f where $\bar{e}$ is the identity of $\bar{G}$. Then:

(a) H is a closed normal subgroup of G and $\pi : G \to G/H$ is a continuous homomorphism;

(b) there is a continuous monomorphism $g : G/H \to \bar{G}$ such that $f = g \circ \pi$;

(c) let H and N be closed normal topological subgroups of G such that $N \subset H$. Then G/H is topologically isomorphic to $(G/N)/(H/N)$.

Example (1) Let R be the additive group of the reals with the usual metric topology and let Z be the additive subgroup of the integers. Then Z is closed in R and the quotient group is topologically isomorphic to the multiplicative group of complex numbers of absolute value 1 denoted by T^1 or S^1; that is, the one-dimensional torus. Then R/Z is frequently called the (one-dimensional) **torus group** or the **group of reals modulo 1**. The above isomorphism uses the fact that the map $f: R \to S^1 : x \to e^{2\pi ix}$ is a continuous epimorphism with kernel Z. Thus, from Theorem 3.16, S^1 is topologically isomorphic to R/Z.

Exercise (2) Generalize the above example by finding an explicit homomorphism $f: R^n \to T^n$ of the topological groups. What is the kernel of this map?

Definition 3.17 Let M be a Hausdorff topological space and let G be a topological group. Then:

(a) G **operates** on M if there is a surjection

$$G \times M \to M : (g, p) \to g \cdot p$$

such that $(g_1 g_2) \cdot p = g_1 \cdot (g_2 \cdot p)$ and $e \cdot p = p$ for all $g_1, g_2 \in G$ and $p \in M$ where e is the identity of G.

(b) G **operates transitively** on M if for every $p, q \in M$, there exists $g \in G$ such that $g \cdot p = q$.

(c) G **operates continuously** on M if the map $G \times M \to M : (g, p) \to g \cdot p$ is continuous.

(d) G is called a **topological transformation group** on M if G operates continuously on M. [Note that for each $g \in G$, the map $\tau(g) : M \to M : p \to g \cdot p$ is a homeomorphism.]

(e) G is **effective** if $a \cdot p = p$ for all $p \in M$ implies $a = e$.

(f) Let p be fixed in M. Then $G(p) = \{g \in G : g \cdot p = p\}$ is a group called the **isotropy** subgroup of G at p or **fixed point** subgroup at p. The set $G \cdot p = \{g \cdot p \in M : g \in G\}$ is called an **orbit** under G.

Exercise (3) If G acts transitively on M, then for given $p, q \in M$ the isotropy subgroups $G(p)$ and $G(q)$ are conjugate in G.

Example (2) Let G be a topological group and let H be a closed subgroup. Then the space $M = G/H$ is a Hausdorff space according to Theorem 3.11 and G operates continuously on M by the map $G \times M \to M : (g, xH) \to (gx)H$. For each $g \in G$ the map $\tau(g) : M \to M : xH \to gxH$ is a homeomorphism, and using this, we see G acts transitively on M. The coset space $M = G/H$ is called a **homogeneous space**.

Exercise (4) Show that G is effective on G/H if and only if H contains no proper normal subgroup of G.

Theorem 3.18 Let M be a Hausdorff space and let G be a transitive topological transformation group operating on M. Let p be some (fixed) point in M and let $G(p)$ be the isotropy group at p. Then $G(p)$ is a closed subgroup of G, and the map $f: G \to G \cdot p : a \to a \cdot p$ induces a continuous bijection $\bar{f}: G/G(p) \to M$ such that $\bar{f} \circ \pi = f$; that is, the accompanying diagram is commutative.

$$\begin{tikzcd} G \arrow[rr, "f"] \arrow[dr, "\pi"'] & & M \\ & G/G(p) \arrow[ur, "\bar{f}"'] & \end{tikzcd}$$

PROOF First, since G is transitive, the orbit $G \cdot p$ equals M. Next, since the map $f: G \to M : a \to a \cdot p$ is continuous and $\{p\}$ is closed in the Hausdorff space M, we have $G(p) = f^{-1}(p)$ is closed.

Now note that for $a \in G$ and $h \in G(p)$ we have $f(ah) = ah \cdot p = a \cdot (h \cdot p) = a \cdot p = f(a)$. Thus the map

$$\bar{f}: G/G(p) \to M : aG(p) \to f(a)$$

is actually a well-defined function, for if $aG(p) = bG(p)$, then $b = ah$ for some $h \in G(p)$ and therefore $f(b) = f(ah) = f(a)$. Next $\bar{f}$ is bijective since $f(a) = f(b)$ implies $a^{-1}b \cdot p = p$ and therefore $a^{-1}b \in G(p)$; that is, $aG(p) = bG(p)$. Finally, since we clearly have $f = \bar{f} \circ \pi$ and since $f: G \to M$ is continuous, we have from Theorem 3.11 that $\bar{f}$ is continuous.

Corollary 3.19 If $\bar{f}: G/G(p) \to M$ is open or if $G/G(p)$ is compact, then $\bar{f}$ is a homeomorphism; that is, M is a homogeneous space.

PROOF If $\bar{f}$ is open, then by definition of continuity $\bar{f}^{-1}$ is continuous. If $G/G(p)$ is compact, then we use the following general results: Let S be a compact space and let T be a Hausdorff space. Then any continuous bijection $g : S \to T$ is a homeomorphism [Singer and Thorpe, 1967, p. 24].

REMARK (2) If G is compact, then M is a homogeneous space, for in this case $G/G(p)$ is compact since $\pi : G \to G/G(p)$ is continuous. Note there is almost a converse statement: If H is a closed compact subgroup of G such that G/H is compact, then G is compact [Chevalley, 1946, p. 31; Hochschild, 1965, p. 8].

Corollary 3.20 If G is a locally compact group with countable basis and if M is a locally compact Hausdorff space, then $\bar{f}$ is a homeomorphism of $G/G(p)$ onto M.

The proof of this can be found in the work of Helgason [1962, p. 111].

Example (3) For $n \geq 2$, let $S^{n-1} = \{x \in R^n : \|x\| = 1\}$ be the unit sphere where $\|x\|^2 = B(x, x)$ is the usual inner product on R^n. Let $G = O(n) = \{A \in GL(n, R) : B(Ax, Ax) = B(x, x) \text{ all } x \in R^n\}$ and let $p = (1, 0, \ldots, 0) \in S^{n-1}$.

Then $O(n)$ operates continuously and transitively on S^{n-1} by the map

$$O(n) \times S^{n-1} \to S^{n-1} : (A, x) \to Ax.$$

From this we see that $Ap = p$ if and only if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & B \end{bmatrix}$$

where $B \in O(n-1)$ so that we obtain $G(p) = O(n-1)$. Thus using $O(n)$ is compact [exercise (5) below], its continuous image $O(n)/O(n-1)$ is compact and therefore $O(n)/O(n-1)$ is homeomorphic to S^{n-1}.

In the work of Chevalley [1946, p. 32] it is similarly shown that some of the other groups discussed in Section 2.3 also yield homogeneous spaces which are homeomorphic to spheres. For example, if $n \geq 2$, then we also have $SO(n)/SO(n-1)$ is homeomorphic to S^{n-1}.

Exercise (5) Show $O(n)$ is compact possibly as follows. First using the above representation of $O(n)$ in terms of B, show $O(n)$ is a closed subset of $\mathrm{End}(R^n)$. Next using $AA^* = I$, show $O(n)$ is bounded; thus it is compact.

3. Connected Groups

In this section we shall show that much of the topology and many other relations are determined by the connected component of the identity element of a topological group.

Definition 3.21 Let M be a topological space and let $p \in M$. Then p is contained in a unique maximal connected subset $C(p)$. This set $C(p)$ is closed and is called the **connected component** of p. For $M = G$ a topological group, the connected component of the identity $e \in G$ is called the **identity component** of G and is denoted by G_0.

Theorem 3.22 Let G be a topological group, and let G_0 be the identity component. Then:

(a) G_0 is a closed normal topological subgroup of G and the connected component $C(a)$ of $a \in G$ equals aG_0;

(b) If G is locally connected (that is, if every point $a \in G$ has a connected neighborhood), then G/G_0 is discrete.

PROOF First, since G_0 is the component $C(e)$, we have G_0 is closed. Next let $a \in G_0$. Then since multiplication is continuous, $a^{-1}G_0$ is connected and contains e. Thus $a^{-1}G_0$ is a connected set containing e, and since G_0 is the maximal connected set containing e, we have $a^{-1}G_0 \subset G_0$ so that G_0 is a subgroup. Also for any $x \in G$ we see $x^{-1}G_0 x$ contains e and is connected. Therefore $x^{-1}G_0 x \subset G_0$ so that G_0 is normal.

Next since $x \to ax$ is a homeomorphism of G, then aG_0 is connected and contains a so that $aG_0 \subset C(a)$. Also $a^{-1}C(a)$ is connected and contains e so that $a^{-1}C(a) \subset G_0$; that is, $C(a) = aG_0$.

For (b), let G be locally connected so that there is a connected neighborhood U of e in G. Then since π is an open map, $\pi(U)$ is a neighborhood of eG_0 in G/G_0. However, since U is connected we have $U \subset G_0$ so that $\pi(U) = \{eG_0\}$. Thus $\{eG_0\}$ is open so that G/G_0 is discrete.

Proposition 3.23 Let G be a topological group, let G_0 be the identity component, and let U be any open neighborhood of $e \in G$.

(a) If U is a symmetric neighborhood, then $H = \bigcup_{k=1}^{\infty} U^k$ is an open and closed subgroup of G. If U is connected, so is H.

(b) $G_0 = (\bigcup_{k=1}^{\infty} U^k) \cap G_0$.

(c) If G is connected, then $G = \bigcup_{k=1}^{\infty} U^k$. Thus any open neighborhood of e is a set of generators of a connected topological group as an abstract group.

PROOF (a) If U is symmetric, then for $x \in U^m$, $y \in U^n$, we have $xy \in U^{m+n}$ and $x^{-1} \in (U^{-1})^m = U^m$ so that H is a subgroup. Next since U is open, $U^2 = \bigcup \{aU : a \in U\}$ is open and by induction U^k is open. Thus H is an open subgroup. However, from remark (1) of Section 3.2, H is also closed. If U is connected, then so is each U^k and therefore H is connected (using $e \in U^k$).

(b) Let V be a symmetric neighborhood of e such that $V \subset U$ and let $W = V \cap G_0$. Then W is a symmetric neighborhood of e in G_0 and $H = \bigcup W^k \subset \bigcup V^k \cap G_0$. However, H is a nonempty open and closed subgroup of G_0 and since G_0 is connected we have $G_0 = H$. Since $\bigcup V^k \subset \bigcup U^k$, we obtain the result.

(c) Part (c) follows from (b) since $G = G_0$.

Definition 3.24 Let G be a topological group. Then the **center** C of G equals $\{x \in G : xa = ax \text{ for all } a \in G\}$. The center is a normal subgroup of G and is also denoted by $Z(G)$.

Proposition 3.25 Let G be a connected topological group and let H be a discrete normal topological subgroup of G. Then $H \subset C$, the center of G.

Proof Let $a \in H$. Then the map $G \to H : x \to x^{-1}ax$ is continuous. However, since G is connected and H is discrete, the image is a single point $p \in H$. In particular $p = e^{-1}ae = a$ so that $x^{-1}ax = a$ for all $x \in G$; that is, $H \subset C$.

Proposition 3.26 Let G be a topological group and let H be a closed topological subgroup such that H is connected and G/H is connected. Then G is connected.

Proof Let H and G/H be connected and assume $G = U \cup V$ where U and V are nonempty open sets. The open map $\pi : G \to G/H$ maps U and V onto open sets $U_1 = UH$ and $V_1 = VH$ in G/H. Since $G = U \cup V$, we have $G/H = U_1 \cup V_1$, and since G/H is connected, there exists $aH \in U_1 \cap V_1 = UH \cap VH$. Thus $aH \in UH$ yields $h \in H$ with $ah = u \in U$. Thus $aH \cap U$ is not empty. Similarly $aH \cap V$ is not empty. However, since $G = U \cup V$, we have $aH = (aH \cap U) \cup (aH \cap V)$, and since H is connected and aH is homeomorphic to H, we have aH is connected. Thus $(aH \cap U) \cap (aH \cap V)$ is not empty. This implies $U \cap V$ is not empty so that G is connected.

Definition 3.27 A topological space M is **locally Euclidean** of dimension m if each point $p \in M$ has a neighborhood which is homeomorphic to an open set in R^m. Note that an open subset of R^m cannot be homeomorphic to an open subset of R^n if $m \neq n$.

Examples (1) The torus $T^1 = R/Z$ is connected and R is connected, but Z is not connected. Thus the connectedness of a subgroup H cannot be deduced from that of G and G/H. Also note T^1 is locally Euclidean of dimension 1.

(2) Using the result that the sphere S^{n-1} is connected and from Section 3.2 that $SO(n)/SO(n-1)$ is homeomorphic to S^{n-1}, we shall show $SO(n)$ is connected. Consequently, since $SO(n)$ is actually a C^∞-manifold, we have from Proposition 2.15 that $SO(n)$ is path connected. First note that $SO(1)$ is just the identity linear transformation I. Assume $SO(n-1)$ is connected. Then since $SO(n)/SO(n-1)$ is connected (because it is homeomorphic to S^{n-1}) we have by Proposition 3.26 that $SO(n)$ is connected. Also we previously noted that $O(n)$ is not connected, but note that $SO(n)$ is the identity component of $O(n)$ and the order of $O(n)/SO(n)$ is 2. Thus $O(n)$ has two components, one which consists of matrices of determinant -1 and the other is $SO(n)$. In a similar manner, it is shown by Chevalley [1946, p. 36] that various other matrix groups are connected.

CHAPTER 4

LIE GROUPS

We now discuss some elementary results of Lie groups which can be easily done without introducing the Lie algebra. First we see that a Lie group is a topological group which is an analytic manifold so that there is compatibility between the topological, manifold, and group structures. Next we give results which tell when a topological group is a Lie group and when a local group generates a Lie group. Finally we discuss Lie subgroups and when an abstract subgroup can be considered as a Lie subgroup.

1. Basic Structures

In this section we give the basic definitions for a Lie group and show how the analytic structure of a Lie group is uniquely determined.

Definition 4.1 A **Lie group** is a set G such that:

(a) G is a group;
(b) G is an analytic manifold;
(c) the group multiplication in (a) of the product manifold

$$\mu : G \times G \to G : (x, y) \to xy$$

and the group inversion operation in (a)

$$\iota : G \to G : x \to x^{-1}$$

are analytic functions relative to the structure in (b).

REMARKS (1) A Lie group is, in particular, a topological group relative to the topology induced by its analytic structure, and the question arises when is a topological group actually a Lie group. This is discussed by Montgomery and Zippin [1955, p. 184], and among the many results is the following: A connected locally Euclidean topological group is isomorphic to a Lie group.

(2) The fact that $\iota : G \to G : x \to x^{-1}$ is analytic follows from (a) and (b) and $\mu : G \times G \to G$ is analytic by using the implicit function theorem. Briefly, let (U, x) be an analytic chart at e in G and for u, v in a suitable open nucleus $V \subset U$, we have that $x_i \circ \mu(u, v) \equiv \mu_i(u, v)$ defines an analytic function with $\mu_i(v, e) = \mu_i(e, v) = v_i$. Thus $D_1(x \circ \mu)(e, e) = (\partial\mu_i/\partial x_j(e, e)) = (\delta_{ij})$. By a variation of the implicit function theorem, the equation $x \circ \mu(z, v) = x(e)$ has a solution $z = \theta(v)$ in some neighborhood of e where θ is actually analytic; that is, $\theta_i = x_i \circ \theta$ are analytic. However, $zv = e$ has the solution $z = v^{-1}$ so that the map $v \to v^{-1}$ is analytic at e. Now by using the analyticity of the left and right translations (from $\mu : G \times G \to G$ is analytic), we have ι is analytic on all of G.

Examples (1) The matrix groups $GL(V)$, $SO(n, R)$, etc. previously considered are Lie groups. Also the torus T^1 and more generally T^n is a Lie group. For this, we use $x \to e^{2\pi i x}$ is analytic and so is the multiplication $\mu(e^{2\pi i x}, e^{2\pi i y}) = e^{2\pi i(x+y)}$. Next use the fact that if $G_1, \ldots, G_n$ are Lie groups, then $G_1 \times \cdots \times G_n$ with the product group and analytic structure is again a Lie group which we leave as an exercise.

(2) If G is a discrete topological group, then e has the open neighborhood $\{e\}$ which is homeomorphic to $R^0 = \{0\}$; that is, a discrete topological group can be considered as a zero-dimensional Lie group and conversely.

(3) Let R denote the manifold of the real numbers with the usual coordinate $u : R \to R : t \to t$ and define

$$\mu : R \times R \to R : (x, y) \to (x^3 + y^3)^{1/3}.$$

Then (R, μ) is a topological group but not a Lie group, since μ is not analytic at $(0, 0)$ relative to the above coordinate.

Exercise (1) Show that R with the above multiplication μ becomes a Lie group relative to the analytic structure on R given by $v : R \to R : t \to t^3$.

We now consider the existence and uniqueness of an analytic structure determined by a nucleus.

Proposition 4.2 Let G be a connected topological group with multiplication $\mu(s, t) = st$, let (U, x) be a chart at e in G, and let $V \subset U$ be an open nucleus such that $\mu(V, V) \subset U$. If the function $x \circ \mu$ is analytic on $V \times V$, then there exists a unique analytic structure $\mathscr{A}$ on G which makes G into a Lie group $\mathscr{G}$ such that $(W, x) \in \mathscr{A}$ where W is a suitable open nucleus contained in U. In this case the topology induced by the analytic structure on $\mathscr{G}$ equals the original topology of G.

PROOF First we define the analytic structure $\mathscr{A}$ on G. Let the nucleus V be as above and let W be an open nucleus such that $WW^{-1} \subset V$. Then since $e \in W$ we note $W \subset V \subset U$ and for any $a \in G$ we have aW is a neighborhood of a in G. Now (W, x) defines an analytic chart at $e \in G$ and we now define coordinates on the neighborhood aW by

$$y : aW \to x(W) : au \to x(u).$$

Thus since $x(W)$ is open in R^m, we obtain a chart (aW, y), where for $p \in aW$ we have $y(p) = x \circ L(a)^{-1}(p)$. Furthermore since $x : W \to R^m$ is analytic so is $y : aW \to R^m$.

We shall next show that these charts aW with $a \in G$ are actually analytically related. Thus they form an analytic atlas $\mathscr{B}$ which covers G, and we obtain an analytic structure by taking the maximal atlas $\mathscr{A}$ which contains $\mathscr{B}$. So suppose (aW, y) and $(bW, \bar{y})$ are charts, and let $p \in aW \cap bW$. Then we must show $y \circ \bar{y}^{-1}$ and $\bar{y} \circ y^{-1}$ are analytic. We have from $p = \mu(a, u) = au \in aW$ and $p = bv \in bW$ that $b^{-1}a = vu^{-1} \in WW^{-1} \subset V$. Therefore the map $L(b^{-1}a) : V \to U : z \to \mu(b^{-1}a, z)$ is analytic on V by hypothesis. Next we have from $y = x \circ L(a)^{-1}$ and $\bar{y} = x \circ L(b)^{-1}$ that

$$\bar{y} \circ y^{-1} = x \circ L(b)^{-1} \circ L(a) \circ x^{-1} = x \circ L(b^{-1}a) \circ x^{-1}$$

is an analytic function from $x(W) \subset R^m$ into $x(U) \subset R^m$. Similarly $y \circ \bar{y}^{-1}$ is analytic.

By definition, the chart (W, x) is in the analytic structure $\mathscr{A}$ defined above. Let $\mathscr{G}$ denote the analytic manifold G with the analytic structure $\mathscr{A}$. Then by definition of $\mathscr{A}$ the map $L(a^{-1})$ is an analytic diffeomorphism which maps a neighborhood of $a \in \mathscr{G}$ onto a neighborhood of $e \in \mathscr{G}$. Consequently, since the multiplication μ is analytic at $(e, e) \in \mathscr{G} \times \mathscr{G}$, it is analytic on all of $\mathscr{G} \times \mathscr{G}$; that is, $\mathscr{G}$ is a Lie group; (see exercise (2) below).

The analytic structure $\mathscr{A}$ on the Lie group $\mathscr{G}$ is unique, since it is completely determined by the given chart (U, x) at $e \in G$. Thus if another open nucleus $W_1 \subset U$ determines an analytic structure $\mathscr{A}_1$ on G using the coordinate map x; then by considering $W \cap W_1$, we see that the identity $i : \mathscr{G} \to \mathscr{G}_1$ and its inverse $i : \mathscr{G}_1 \to \mathscr{G}$ are both analytic, where $\mathscr{G}_1$ is the group corresponding to the analytic structure $\mathscr{A}_1$ on G. Finally the map $i : G \to \mathscr{G}$ is a homeomorphism at e and therefore everywhere; that is, the topology induced by the analytic structure on $\mathscr{G}$ equals the original topology on G.

Definition 4.3 Let G and G_1 be Lie groups. Then the map $\phi : G \to G_1$ is an **analytic homomorphism** if ϕ is an analytic mapping which is a homomorphism of the groups. Thus ϕ is an **analytic isomorphism** if ϕ is an analytic homomorphism such that ϕ^{-1} exists and ϕ and ϕ^{-1} are both analytic (diffeomorphisms).

REMARKS (3) From Proposition 4.2, we see that in a topological group a chart at $e \in G$ with local analytic multiplication determines a unique Lie group structure. Now suppose that two charts at e are given such that the multiplication μ is analytic relative to these charts. Then possibly two distinct analytic structures $\mathscr{A}_1$ and $\mathscr{A}_2$ can be determined which give rise to two Lie groups $\mathscr{G}_1$ and $\mathscr{G}_2$. However, we shall show that this is not possible by eventually showing continuous isomorphisms of Lie groups are analytic isomorphisms. Thus the analytic structure is completely determined by the topology.

Exercise (2) Show in detail that the multiplication μ in the proof of Proposition 4.2 is analytic at any point $(a, b) \in \mathscr{G} \times \mathscr{G}$ possibly as follows. First show inversion $\iota : \mathscr{G} \to \mathscr{G} : x \to x^{-1}$ is analytic at e. Using this and $\mathscr{G}$ is connected, show for any $a \in \mathscr{G}$ that $R(a) = \iota \circ L(a^{-1}) \circ \iota$ is analytic at e. Next observe the map $\alpha \times \beta : \mathscr{G} \times \mathscr{G} \to \mathscr{G} \times \mathscr{G} : (x, y) \to (a^{-1}x, yb^{-1})$ is analytic and use the factorization $\mu = R(b) \circ L(a) \circ \mu \circ (\alpha \times \beta)$ and μ is analytic at (e, e) to show μ is analytic at (a, b).

2. Local Lie Groups

We now consider local groups which are manifolds and use these to also determine Lie groups. Since we are considering locally Euclidean groups, we shall henceforth assume nuclei to be connected.

Definition 4.4 A **local Lie group** is a set B such that:

(a) B is a connected analytic manifold;

(b) B is a local group relative to the topology induced from the analytic structure in (a);

(c) there is an open germ U of the local group B such that for the multiplication function μ in B, the map $\mu : U \times U \to B : (x, y) \to \mu(x, y)$ is analytic.

REMARK (1) As discussed for Lie groups, a local Lie group defines an analytic inverse operation $\iota : W \to W : x \to x^{-1}$ defined on a suitable open neighborhood W of e in U.

(2) An analytic vector field yields a local one-parameter group ϕ on an analytic manifold M and ϕ is a local Lie group.

(3) An open nucleus of a Lie group is a local Lie group.

Definition 4.5 (a) Two local Lie groups, B_1 and B_2, are **locally analytically isomorphic** if there exists a local isomorphism $f\colon B_1 \to B_2$ such that f and f^{-1} are analytic.

(b) Two Lie groups G and G' are **locally isomorphic** if they have open nuclei which as local Lie groups are locally analytically isomorphic.

Example (1) An open neighborhood B_1 of e in the torus T^1 and an open neighborhood B_2 of $0 \in R$ yield locally analytically isomorphic local Lie groups, but T^1 and R are not isomorphic Lie groups.

Related to Proposition 4.2 is the following result [Cohn, 1957]:

Proposition 4.6 Let G be an abstract group and let B be a subset of G such that:

(a) B generates G as a group;
(b) B is a (connected) local Lie group relative to the multiplication in G.

Then there is defined on G exactly one analytic structure $\mathscr{A}$ which makes G into a connected Lie group $\mathscr{G}$ so that:

(i) the group structure of $\mathscr{G}$ is the group structure of G;
(ii) for some open nucleus U of B with coordinate map x, the chart $(U, x) \in \mathscr{A}$.

Corollary 4.7 Let G be a connected topological group which is a (topological) manifold and let $\phi : G \times G \to G : (x, y) \to xy^{-1}$ $[= \mu(x, y^{-1})]$. Let U be an open neighborhood of e which has a given analytic structure so that the map $\phi^{-1}(U) \cap (U \times U) \to U : (x, y) \to xy^{-1}$ is analytic. Then there exists a unique analytic structure $\mathscr{A}$ on G relative to which G becomes a Lie group $\mathscr{G}$ such that the topology of G equals the topology of $\mathscr{G}$ and the analytic structure of $\mathscr{G}$ restricted to a suitable open nucleus $V \subset U$ is equivalent to the given analytic structure on U.

PROOF (of Proposition 4.6) First we define the topology on G as follows. Let $\mathscr{V}$ be the family of nuclei of B, and let $\mathscr{W} = \{W \subset G : W \cap B \in \mathscr{V}\}$. Then $\mathscr{W}$ is nonempty, since $\mathscr{V} \subset \mathscr{W}$ and $\mathscr{W}$ satisfies the conditions of Proposition 3.6 as follows. Conditions (a) and (b) are clear. For (c), let $W_1 \in \mathscr{W}$ so that $W_1 \cap B \in \mathscr{V}$. Then there exists $V \in \mathscr{V} \subset \mathscr{W}$ with $VV^{-1} \subset W_1 \cap B \subset W_1$. For (d), let $W \in \mathscr{W}$ and $a \in G$. Then since B generates G as a group, we can write $a = a_1 \cdots a_n$ where a_i or a_i^{-1} is in B. Now since $a_i^{-1} e a_i = e$ and W is a neighborhood of e, we have by continuity of the multiplication that there exists $V_n \in \mathscr{V}$ with $a_n^{-1} V_n a_n \subset W$. Similarly there exists $V_{n-1} \in \mathscr{V}$ such that $a_{n-1}^{-1} V_{n-1} a_{n-1} \subset V_n$, and by induction, there exists $V_k \in \mathscr{V}$ such that $a_k^{-1} V_k a_k \subset V_{k+1}$ for $k = 1, \ldots, n-1$. Thus

$$\begin{aligned} a^{-1} V_1 a &= a_n^{-1} \cdots a_1^{-1} V_1 a_1 \cdots a_n \\ &\subset a_n^{-1} \cdots a_2^{-1} V_2 a_2 \cdots a_n \\ &\;\vdots \\ &\subset a_n^{-1} V_n a_n \subset W. \end{aligned}$$

Thus $V_1 \subset aWa^{-1}$, and since $V_1 \in \mathscr{V} \subset \mathscr{W}$, we have by condition (b) applied to $\mathscr{W}$ that $aWa^{-1} \in \mathscr{W}$. Condition (e) holds for $\mathscr{W}$, since it holds for $\mathscr{V}$. Thus by Proposition 3.6, G becomes a topological group with $\mathscr{W}$ as a family of of nuclei.

Now the topology defined on G by $\mathscr{W}$ restricts to the original manifold topology given on B. For $\mathscr{V} \subset \mathscr{W}$ and, conversely, any $W \in \mathscr{W}$ contains a set $W \cap B$ which is in $\mathscr{V}$. Therefore W is a nucleus of G. Thus the topological group operations on $(G, \mathscr{W})$ are actually analytic near e so that by Proposition 4.2, G can now be defined to be a Lie group $\mathscr{G}$ [noting by Proposition 3.23(a), B generates G so that G is connected].

3. Lie Subgroups

We shall now define the concept of a Lie subgroup of a Lie group G and note that this concept differs from a topological subgroup because a Lie subgroup need not have the induced topology.

Definition 4.8 Let G be a Lie group and let H be a Lie group. Then H is a **Lie subgroup** of G if H is an analytic submanifold of G and if H is a subgroup of G.

Examples (1) The torus T^2 is a Lie group when regarded as a product group $T^1 \times T^1$ and has as a submanifold the "irrational wrap around" curve as discussed in Section 2.3 (see Fig. 4.1). This curve is given by $f(t) = (\exp 2\pi iat, \exp 2\pi ibt)$, where $a/b = \alpha$ is irrational. As we saw, this curve is a one-dimensional submanifold which is dense in T^2, and since $f(s+t) = f(s)f(t)$ in T^2, it is therefore a Lie subgroup which is not closed. This Lie

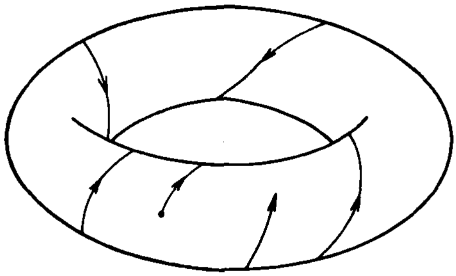


Fig. 4.1.

subgroup does not have the induced topology since there are points on the curve which are arbitrarily close in the topology induced from T^2 but are arbitrarily far apart in the topology of the curve.

(2) The integers Z are a zero-dimensional Lie subgroup of R.

Exercise (1) Every discrete subgroup of a Lie group is a closed Lie subgroup (*Hint:* $\{e\}$ is a neighborhood of e.)

We shall now give various criterion for a subgroup to be a Lie subgroup and we need the following result [Helgason, 1962, p. 78].

Lemma 4.9 Let M and N be C^∞ (analytic) manifolds, and let $f\colon M \to N$ be a C^∞ (analytic) mapping such that $f(M)$ is contained in a submanifold P. If the map $f\colon M \to P$ is continuous, then this map is also C^∞ (analytic).

PROOF We shall show this as follows. Let the accompanying diagram be

$$\begin{tikzcd} M \arrow[rr, "f"] \arrow[dr, "F"'] & & N \\ & P \arrow[ur, "i"'] & \end{tikzcd}$$

commutative, where f is C^∞, F continuous, and i an immersion (since P is a submanifold). Then by the remarks following Proposition 2.23, for each $p \in P$ there is a neighborhood U of p, and a neighborhood V of $i(p) \in N$ and a C^∞-map $g : V \to U$ so that $g \circ i = \text{identity}\,|\,U$. Thus since $f = i \circ F$, we have locally that

$$F = \text{identity} \circ F = g \circ (i \circ F) = g \circ f,$$

and since the right side is a composition of C^∞-functions, F is C^∞. In particular, letting $F = f$ and i be the identity; that is, letting P be a submanifold, we obtain the result.

Proposition 4.10 Let G be a Lie group and let H be a submanifold of G which is also an abstract subgroup of G. If H is also a topological group (relative to the topology induced from its analytic structure), then H is a Lie subgroup of G.

PROOF It suffices to show H is a Lie group and this follows from the preceding lemma. The mapping $f: G \times G \to G : (x, y) \to xy^{-1}$ is analytic and its restriction $f_H : H \times H \to G$ is also analytic. Now since H is a topological group, the map $f_H : H \times H \to H$ is continuous. Thus by Lemma 4.9, f_H is analytic so that H is a Lie group.

The next result follows from previous facts.

Proposition 4.11 Let G be a Lie group and let H be a connected topological subgroup of G. Then there is at most one analytic structure $\mathscr{A}(H)$ on H which makes H into a Lie subgroup of G.

We now give some computational results which determine Lie subgroups.

Proposition 4.12 Let H be a Lie group which is an abstract subgroup of the Lie group G. Assume at the identity $e \in G$ there exist an analytic chart (U, x) in G and an analytic chart (V, y) at e in H such that $x_i | H = y_i$ and $(\partial y_i/\partial x_j(e))$ has rank equal to the dimension of H. Then H is a Lie subgroup of G.

PROOF We first translate the charts at e to any point $a \in H$ by the analytic diffeomorphisms $L(a)$ (of H and G) so that we can now apply Corollary 2.12 to obtain H is a submanifold.

This result can also be stated in terms of local Lie groups which generate a subgroup.

Corollary 4.13 Let G be a Lie group and let B be a local Lie group relative to the group operations in G. If there exist charts (U, x) at e in G and (V, y) at e in B such that $x_i | B = y_i$ and rank $(\partial y_i/\partial x_j(e)) = \dim B$, then the subgroup H generated by B is a connected Lie subgroup of G.

We now note that the topological and manifold structure is mostly in the identity component. This will also become more evident when the Lie algebras are also taken into consideration.

Proposition 4.14 Let G be a Lie group and let G_0 be the identity component of G (as a topological group). Then:

(a) G_0 is an open normal Lie subgroup of G;
(b) $T(G, e) = T(G_0, e)$ and therefore $\dim G = \dim G_0$;
(c) G/G_0 is a Lie group which is discrete.

PROOF Since G is locally Euclidean, it has a connected open neighborhood U of e in G and from Proposition 3.23, U generates a connected subgroup H of G. Since G_0 is the identity component, we have by maximality that $G_0 \supset H$. However, H contains the neighborhood U of e in G_0 and G_0 is connected. Therefore $G_0 = H$. Now G_0 contains the neighborhood U of e and U is open in G so that any $a \in G_0$ is in the open neighborhood aU. Thus G_0 is open in G. This means G_0 is an open submanifold of G so that $\dim G = \dim G_0$ and $T(G, e) = T(G_0, e)$. Also by Theorem 3.22, G/G_0 is discrete.

REMARKS (1) If G is a connected Lie group and H a proper Lie subgroup, then $\dim H < \dim G$ for otherwise H contains an open nucleus of G which generates G; that is, $G = H$.

(2) We shall show later that if G is a Lie group and if H is a closed subgroup of G, then H is a Lie subgroup of G. Thus the previously discussed subgroups $O(n)$, $SL(n)$, and $Sp(n)$ are all closed Lie subgroups of $GL(n, R)$.

(3) We shall consider later normal Lie subgroups when we discuss homomorphisms.

CHAPTER 5

THE LIE ALGEBRA OF A LIE GROUP

We have seen from previous examples that the tangent space $T(G, e)$ of a Lie group G can be used to give local information about G. In this chapter we formalize this situation by introducing the set of G-invariant vector fields $\mathscr{L}(G)$ and seeing that it is a vector space which is isomorphic to $T(G, e)$. Also $\mathscr{L}(G)$ is a Lie algebra over R and induces a Lie algebra structure on $T(G, e)$. Using this we define the exponential map $\exp : \mathscr{L}(G) \to G$ in terms of homomorphisms of R into G. The exponential map is a local diffeomorphism $\exp : U_0 \to U_e$ of a suitable neighborhood U_0 of 0 in $\mathscr{L}(G)$ onto a neighborhood U_e of e in G. Using the inverse function $\log : U_e \to U_0$ we define canonical coordinates $(U_e, \log)$ at e in G. Thus by the action of $L(a) : G \to G : x \to ax$ we obtain coordinates at any point $a \in G$.

The exponential map is used to obtain a local representation of the multiplication in G analogous to the results of Section 1.6. Thus for X and Y sufficiently near 0 in $\mathscr{L}(G)$ we can write $\exp X \exp Y = \exp F(X, Y)$ where $F: \mathscr{L}(G) \times \mathscr{L}(G) \to \mathscr{L}(G)$ is analytic at $(0, 0) \in \mathscr{L}(G) \times \mathscr{L}(G)$. We show that the terms $F^k(0, 0)(X, Y)^{(k)}$ of the Taylor's series for F are in the subalgebra of $\mathscr{L}(G)$ generated by X and Y. We briefly discuss the actual formula for $F(X, Y)$ which is known as the Campbell–Hausdorff formula. Finally we show that a continuous homomorphism of Lie groups is analytic. This yields the fact that the analytic structure of a Lie group is uniquely determined by its topology.

1. The Lie Algebra

We now introduce the Lie algebra of a Lie group G in terms of invariant vector fields. Thus the Lie algebra will be determined by the tangent space $T(G, e)$ and the action of G determines the values of the vector fields at any other point in G.

Definition 5.1 An analytic vector field $X \in D(G)$ defined on a Lie group G is called **invariant** if for all $a \in G$

$$[(TL(a))(e)]X(e) = X(a).$$

Thus as in Section 2.7 we have that since $(TL(a))(e) : T(G, e) \to T(G, a)$, then the value $[(TL(a))(e)]X(e)$ actually equals $X(a)$.

Next we note that if X is invariant, then X is $L(a)$-invariant for all $a \in G$; that is, X is actually **G-invariant** or **left invariant** according to Section 2.7. For let $p \in G$, then

$$\begin{aligned} X(L(a)p) = X(ap) &= [(TL(ap))(e)]X(e) \\ &= [T(L(a) \circ L(p))(e)]X(e) \\ &= [TL(a)(p)] \cdot (TL(p)(e))(X(e)) \\ &= TL(a)(p) \cdot X(p) \end{aligned}$$

which gives the result.

Proposition 5.2 Let G be a Lie group, let $X \in T(G, e)$, and let

$$\tilde{X} : G \to T(G) : p \to \tilde{X}(p),$$

where $T(G)$ is the tangent bundle of G with projection map π and $\tilde{X}(p)$ is given by

$$(\tilde{X}f)(p) = X(f \circ L(p))$$

where f is any real-valued analytic function on G. Then $\tilde{X}$ is a G-invariant analytic vector field on G such that $\tilde{X}(e) = X$. Furthermore $\tilde{X}$ is the unique G-invariant vector field on G such that $\tilde{X}(e) = X$. Thus any G-invariant vector field is of the form $\tilde{X}$.

Proof Letting $TL(p) = TL(p)(e)$ we first note that $(\tilde{X}f)(p) = (TL(p)X)(f)$ so that $\tilde{X}(p) \in T(G, p)$ and therefore $(\pi \circ \tilde{X})(p) = p$. Thus $\tilde{X}$ is a vector field on G. Since $\tilde{X}(p) = TL(p)X$ we have $\tilde{X}(e) = X$ and the above computations show $\tilde{X}$ is G-invariant. For the uniqueness we use Proposition 2.34 with

$f = L(a)$ for any $a \in G$ or directly as follows. Let Z be a G-invariant vector field with $Z(e) = X$. Then $Z(p) = TL(p)(e)Z(e) = TL(p)(e)X = \tilde{X}(p)$. Finally we shall show $\tilde{X}$ is analytic and derive another formula for it. Thus let $\alpha : I \to G : t \to \alpha(t)$ be an analytic curve on an interval I containing $0 \in R$ so that $\dot{\alpha}(0) = X\ [= \tilde{X}(e)]$ and $\alpha(0) = e$. Then analogous to the results in Section 2.7 we use the results on curves in Section 2.5 to obtain

$$\begin{aligned}(\tilde{X}f)(p) &= X(f \circ L(p)) \\ &= d/dt(0)(f \circ L(p) \circ \alpha) = d/dt[f(p\alpha(t))]_{t=0} \qquad (*)\end{aligned}$$

where $p\alpha(t)$ is the analytic product in G. Thus since f, α, and the multiplication in G are analytic we have $\tilde{X}f$ is an analytic function; that is, $\tilde{X}$ is analytic.

Let $\mathscr{L}(G)$ denote the set of G-invariant vector fields on G. Then from the above result we see that $\mathscr{L}(G)$ consists of all vectors of the form $\tilde{X}$ for $X \in T(G, e)$. From $\tilde{X}(p) = TL(p)(e)X$ and $TL(p)(e)$ being injective we obtain the following.

Corollary 5.3 The map $\phi : \mathscr{L}(G) \to T(G, e) : \tilde{X} \to X$ is a vector space isomorphism. In particular, the dimension of $\mathscr{L}(G)$ over R equals the dimension of G and is finite.

Corollary 5.4 $\mathscr{L}(G)$ is a Lie algebra relative to the bracket operation $[\tilde{X}, \tilde{Y}] = \tilde{X}\tilde{Y} - \tilde{Y}\tilde{X}$.

PROOF This follows from Proposition 2.33.

Definition 5.5 (a) The **Lie algebra of a Lie group** G is the Lie algebra $\mathscr{L}(G)$ of G-invariant vector fields on G.

(b) The Lie algebra g with product $[\ ]_g$ is **homomorphic** to the Lie algebra h with product $[\ ]_h$ if there is a vector space homomorphism $\phi : g \to h$ such that $\phi[XY]_g = [\phi X \phi Y]_h$ for all $X, Y \in g$. If ϕ is a vector space isomorphism, then g and h are **isomorphic** Lie algebras.

By means of Corollary 5.4 we can make $T(G, e)$ into a Lie algebra as follows. Let $X, Y \in T(G, e)$ and let $\tilde{X}, \tilde{Y} \in \mathscr{L}(G)$ as above. Then define the product $[XY] = [\tilde{X}, \tilde{Y}](e)$ which is in $T(G, e)$ and makes $T(G, e)$ into a Lie algebra. This yields the following.

Corollary 5.6 The map $\phi : \mathscr{L}(G) \to T(G, e) : \tilde{X} \to X$ is a Lie algebra isomorphism.

Frequently the Lie algebra $T(G, e)$ is also called the "Lie algebra of G."

Example (1) Let $G = GL(V)$. Then from Corollary 5.6 we have the map $\phi : \mathscr{L}(G) \to T(G, I)$ is a Lie algebra isomorphism using the product $[XY] = [\tilde{X}, \tilde{Y}](I)$ in $T(G, I)$. However, we also have the Lie algebra $gl(V)$ attached to G and we now show that $\mathscr{L}(G)$ is isomorphic to $gl(V)$ as Lie algebras. Recall from example (3), Section 2.5 that for each $A \in gl(V)$ we defined an element $\bar{A} \in T(G, I)$ by

$$(\bar{A}h) = [Dh(I)]A,$$

h analytic at I. The map $gl(V) \to T(G, I) : A \to \bar{A}$ is a vector space isomorphism. Thus we obtain a vector field $\tilde{\bar{A}}$ in $\mathscr{L}(G)$ and consequently a vector space isomorphism $gl(V) \to \mathscr{L}(G) : A \to \tilde{\bar{A}}$. We now show this is a Lie algebra isomorphism; that is, $[\tilde{\bar{A}}, \tilde{\bar{B}}] = \widetilde{\overline{[A, B]}}$. Usually $T(G, I)$ and $gl(V)$ are considered the same and the overbar is omitted as done before but we shall not do this now. Let $p \in G$, and A, $B \in gl(V)$. Then using $L(p)A = pA$, the product in $\mathrm{End}(V)$, we have for f analytic on G

$$
\begin{aligned}
(\tilde{\bar{A}}\tilde{\bar{B}})(f)(p) &= \tilde{\bar{A}}(\tilde{\bar{B}}(f))(p) \\
&= \bar{A}(\tilde{\bar{B}}(f) \circ L(p)), && \text{definition of } \tilde{X} \\
&= [D(\tilde{\bar{B}}(f) \circ L(p))(I)]A, && \text{definition of } \bar{A} \\
&= [D(\tilde{\bar{B}}(f))(p) \circ D(L(p))(I)]A, && \text{chain rule} \\
&= [D(\tilde{\bar{B}}(f))(p)](pA), && L(p) \text{ linear} \\
&= \lim_{t \to 0} \frac{1}{t} [(\tilde{\bar{B}}f)(p + tpA) - (\tilde{\bar{B}}f)(p)] \\
&= \lim_{t \to 0} \frac{1}{t} [\bar{B}(f \circ L(p + tpA)) - \bar{B}(f \circ L(p))] \\
&= \lim_{t \to 0} \frac{1}{t} \{[D(f \circ L(p + tpA))(I)]B - [D(f \circ L(p))(I)]B\} \\
&= \lim_{t \to 0} \frac{1}{t} \{[Df(p + tpA)](pB + tpAB) \\
&\qquad - [Df(p)](pB)\}, && \text{chain rule} \\
&= \lim_{t \to 0} \frac{1}{t} [Df(p + tpA)(pB) - Df(p)(pB)] \\
&\qquad + \lim_{t \to 0} \frac{1}{t} [Df(p + tpA)](tpAB) \\
&= D^2f(p)(pB, pA) + Df(p)(pAB).
\end{aligned}
$$

Interchanging A and B, subtracting the equations, and using the fact that $D^2f(p)$ is symmetric we obtain

$$\begin{aligned}(\tilde{A}\tilde{B} - \tilde{B}\tilde{A})(f)(p) &= Df(p)(pAB - pBA)\\ &= Df(p)(p[A, B])\\ &= [D(f \circ L(p))(I)][A, B]\\ &= \overline{[A, B]}(f \circ L(p))\\ &= (\widetilde{[A, B]})(f)(p)\end{aligned}$$

which proves the result.

Exercises (1) Let G be a Lie group with group multiplication $\mu : G \times G \to G$ and consider the tangent map $(T\mu)(a, b) : T(G, a) \times T(G, b) \to T(G, \mu(a,b))$ as in exercise (3), Section 2.7. From that exercise we obtain vector fields given by $l(\mu, X)(a) = [(T\mu)(a, e)](0, X)$ and $r(\mu, X)(a) = [(T\mu)(e, a)](X, 0)$ for $X \in T(G, e)$ and $0 \in T(G, a)$.

(i) Are the vector fields $l(\mu, X)$ or $r(\mu, X)$ left invariant or right invariant under the action of G; that is, invariant under the set of functions $L(G)$ or $R(G)$?

(ii) If they are invariant, how do they compare with the vector field $\tilde{X}$ given in this section?

(2) Let G and H be Lie groups with Lie algebras g and h. Then as in Section 4.1, the product group $G \times H$ is a Lie group. Show $g \times h$ with the pointwise operations is the Lie algebra of $G \times H$; that is, use the product $[(X_1, Y_1)(X_2, Y_2)] = ([X_1 X_2], [Y_1 Y_2])$.

(3) Let $\mathscr{R}(G)$ be the set of vector fields which are $R(G)$-invariant. Show $\mathscr{R}(G)$ is a Lie algebra. How is it related to $\mathscr{L}(G)$? [Possibly consider $R(a) \circ \iota = \iota \circ L(a^{-1})$ and $[T\iota(a^{-1})]\tilde{X}(a^{-1})$].

(4) Show the vector field $\tilde{X} \in \mathscr{L}(G)$ depends analytically on the parameter $X \in T(G, e)$ (see Section 2.8).

2. The Exponential Map

In this section we generalize the map $\exp : gl(V) \to GL(V)$ to the map $\exp : \mathscr{L}(G) \to G$ which allows us to coordinatize G so that the multiplicative properties of G are nicely translated into properties of the Lie algebra $\mathscr{L}(G)$. First let us recall the function $\exp : gl(V) \to GL(V)$ as discussed in Chapter 1.

We saw that there exist neighborhoods U_0 of 0 in $gl(V)$ and U_I of I in $GL(V)$ so that $\exp : U_0 \to U_I$ is an analytic diffeomorphism. Also for any $X \in gl(V)$ and s, t, $s + t$ near enough $0 \in R$ we have $\exp sX$, $\exp tX$ defined, and $\exp(s + t)X = \exp sX \cdot \exp tX$ in $GL(V)$. Thus identifying X with $\tilde{X}$ in $\mathscr{L}(GL(V))$ we have the fact that $\phi(t) = \exp tX$ is a local one-parameter group on $GL(V)$ determined by X (see Section 2.8). Furthermore we saw that ϕ is actually defined on all of R; that is, $\phi : R \to GL(V)$ is an analytic homomorphism of Lie groups.

Now for an arbitrary Lie group G we would like to start with an arbitrary G-invariant vector field $\tilde{X} \in \mathscr{L}(G)$, form the local one-parameter group $\phi(t)$ determined by $\tilde{X}$, and then extend $\phi(t)$ to a global one-parameter group $\tilde{\phi} : R \to G$. Using this we define $\exp \tilde{X} = \tilde{\phi}(1)$ which is consistent with the results for $GL(V)$. Thus in this section we shall first consider the extension problem so that the exponential map can be defined as described. It should be noted that there are many other possible approaches to this definition depending on the properties one wants to assume [Chevalley, 1946; Helgason, 1962; Loos, 1969].

Lemma 5.7 Let $V = R^m$ be the m-dimensional vector space with the usual Euclidean topology and let B be an open ball with center $0 \in V$. Let G be a topological group and let $\phi : B \to G$ be a continuous local homomorphism of the additive group structure of V; that is, if x, y, $x + y \in B$, then $\phi(x + y) = \phi(x)\phi(y)$ in G. Then there exists a unique continuous homomorphism $f : V \to G$ such that $f | B = \phi$.

PROOF Let $x \in V$ and let p be a positive integer such that $x/n \in B$ for all $n \geq p$. Then we set $f(x) = \phi(x/p)^p$. We now show f is well defined; that is, f is a function. Thus let q be a positive integer such that $x/n \in B$ for all $n \geq q$. Then

$$\begin{aligned} \phi(x/pq)^{pq} &= [\phi(y/q)^q]^p, \qquad \text{where} \quad y = x/p \in B \\ &= [\phi(y/q) \cdots \phi(y/q)]^p \\ &= [\phi(y/q + \cdots + y/q)]^p = \phi(x/p)^p, \end{aligned}$$

so that interchanging p and q we obtain $\phi(x/p)^p = \phi(x/q)^q$. Thus the definition of f is independent of the choice of p so that f is a function. Next, for x, $y \in V$, let p be such that x/p, y/p, and $(x + y)/p \in B$. Then since $\phi(x/p)\phi(y/p) = \phi((x + y)/p) = \phi((y + x)/p) = \phi(y/p)\phi(x/p)$ we obtain

$$\begin{aligned} f(x + y) &= [\phi((x + y)/p)]^p \\ &= [\phi(x/p)\phi(y/p)]^p \\ &= \phi(x/p)^p\phi(y/p)^p = f(x)f(y). \end{aligned}$$

Thus $f : V \to G$ is a homomorphism.

We now show $f|B = \phi$. Thus let $x \in B$ with p as above so that $x/p \in B$ and $x = x/p + \cdots + x/p$ in B. Then using ϕ as a local homomorphism, we obtain

$$\begin{aligned} f(x) &= \phi(x/p)^p \\ &= \phi(x/p) \cdots \phi(x/p) \\ &= \phi(x/p + \cdots x/p) = \phi(x). \end{aligned}$$

Finally f is unique, for if $g : V \to G$ is another continuous homomorphism with $g|B = \phi$, then if $x \in V$ and p such that $x/n \in B$ for all $n \geq p$, we have

$$g(x) = g(p(x/p)) = g(x/p)^p = \phi(x/p)^p = f(x).$$

Corollary 5.8 An analytic local homomorphism α of an interval I about $0 \in R$ into a Lie group G can be extended uniquely to an analytic homomorphism f of R into G.

PROOF Just note that I contains an open interval with center 0 and that the above definition of f is given in terms of analytic operations in R and G.

Theorem 5.9 Let G be a Lie group, let $X \in T(G, e)$, and let $\tilde{X} \in \mathscr{L}(G)$ be the corresponding G-invariant vector field. Then there exists exactly one analytic homomorphism $f: R \to G$ such that $\dot{f}(0) = X$. Thus f is the maximal integral curve of $\tilde{X}$ through $e \in G$; that is, $\dot{f}(t) = \tilde{X}(f(t))$ for all $t \in R$.

PROOF Let $I = (-\varepsilon, \varepsilon)$ and $\alpha : I \to G$ be the integral curve of $\tilde{X}$ such that $\alpha(0) = e$ as discussed in Theorem 2.37. We shall now show that α is an analytic local homomorphism. Let J be a suitable subinterval of I containing $0 \in R$ such that for t fixed in I the analytic maps

$$u : J \to G : s \to \alpha(t + s) \qquad \text{and} \qquad v : J \to G : s \to \alpha(t)\alpha(s)$$

are defined. Then $u(s) = [\alpha \circ \tau(t)](s)$, where $\tau(t) : R \to R : x \to t + x$ and $v(s) = [L(\alpha(t)) \circ \alpha](s)$. Using the results on curves we have

$$\begin{aligned} \dot{u}(s) &= [T(\alpha \circ \tau(t))(d/du)](s) \\ &= (T\alpha)(\tau(t)(s)) \circ [(T\tau(t))(s)](d/du) \\ &= (T\alpha)(t + s)(d/du) \\ &= \dot{\alpha}(t + s) = \tilde{X}(\alpha(t + s)) = \tilde{X}(u(s)), \end{aligned}$$

where we use the chain rule for the second equality, $[T\tau(t)](s)$ is the identity for the third equality, and α is an integral curve for the fifth equality.

Next we have

$$\begin{aligned}
\dot{v}(s) &= [T(L(\alpha(t)) \circ \alpha)(d/du)](s) \\
&= [TL(\alpha(t))](\alpha(s)) \circ (T\alpha)(s)(d/du) \\
&= [(TL(\alpha(t))(\alpha(s))](\dot{\alpha}(s)) \\
&= [TL(\alpha(t))(\alpha(s))]\tilde{X}(\alpha(s)) \\
&= [TL(\alpha(t))\tilde{X}](\alpha(s)) \\
&= [\tilde{X} \circ L(\alpha(t))](\alpha(s)) \\
&= \tilde{X}(\alpha(t)\alpha(s)) = \tilde{X}(v(s)),
\end{aligned}$$

where we use the chain rule for the second equality and $\tilde{X}$ is G-invariant for the sixth equality. Thus we see that $u(s)$ and $v(s)$ are solutions to the differential equation $\dot{z}(s) = \tilde{X}(z(s))$ and satisfy $u(0) = v(0) = \alpha(t)$. Thus by Theorem 2.37 there is a neighborhood J' of $0 \in R$ where $u(s) = v(s)$; that is, there is a suitable neighborhood N of $0 \in R$ so that $\alpha : N \to G$ is an analytic local homomorphism. The theorem now follows from Corollary 5.8 and the following which shows $\dot{f}(t) = \tilde{X}(f(t))$ for all $t \in R$. Thus let $t = t_1 + t_2 \in R$, where $t_1, t_2 \in N$ which is a neighborhood of $0 \in R$ such that f is an integral curve; that is, $f = \alpha$. In particular, $\dot{f}(t_2) = \tilde{X}(f(t_2))$ and $f(t_1 + t_2) = f(t_1)f(t_2)$. Now since $\tilde{X}$ is defined on G, we have from various definitions and the G-invariance of $\tilde{X}$

$$\begin{aligned}
\tilde{X}(f(t)) &= \tilde{X}(f(t_1 + t_2)) \\
&= \tilde{X}(L(f(t_1))f(t_2)) \\
&= [TL(f(t_1))]f(t_2)\tilde{X}(f(t_2)) \\
&= [TL(f(t_1))](f(t_2))\dot{f}(t_2) \\
&= [TL(f(t_1))](f(t_2)) \cdot [(Tf)(d/du)](t_2) \\
&= [T(L(f(t_1)) \circ f)(d/du)](t_2) \\
&= [T(f \circ \tau(t_1))(d/du)](t_2) \\
&= (Tf)(t_1 + t_2) \circ (T\tau(t_1))(t_2)(d/du) \\
&= (Tf)(t)(d/du) = \dot{f}(t).
\end{aligned}$$

Noting that any $t \in R$ can be written as $t = \sum_{i=1}^{n} t_i$, where $t_i \in N$, we use f as a homomorphism and induction on n to obtain the result. Also because the domain of f equals R, f is the unique maximal integral curve through $e \in G$.

Exercise (1) How can the results of Theorem 2.37(b) be formulated in the present context to give the above result directly?

Definition 5.10 Let G be a Lie group and let $\tilde{X} \in \mathscr{L}(G)$ be the vector field corresponding to $X \in T(G, e)$. Let $f_{\tilde{X}}$ denote the unique analytic homomorphism of R into G of Theorem 5.9 such that $\dot{f}_{\tilde{X}}(t) = \tilde{X}(f(t))$. Then we define the **exponential map** exp or $\exp_G$ to be the map with domain $\mathscr{L}(G)$ given by

$$\exp : \mathscr{L}(G) \to G : \tilde{X} \to f_{\tilde{X}}(1).$$

We first note that for all $t \in R$,

$$\exp t\tilde{X} = f_{t\tilde{X}}(1) = f_{\tilde{X}}(t). \qquad (*)$$

The first equality is the definition. To see the second let t be fixed (but arbitrary) and let

$$g : R \to G : s \to f_{\tilde{X}}(ts).$$

Then g is an analytic homomorphism and if $\tau^*(t) : R \to R : x \to tx$ we have $g(s) = [f_{\tilde{X}} \circ \tau^*(t)](s)$. Consequently

$$\begin{aligned}\dot{g}(0) &= [T(f_{\tilde{X}} \circ \tau^*(t))(d/du)](0)\\ &= (Tf_{\tilde{X}})(\tau^*(t)(0)) \circ [T\tau^*(t)](0)(d/du)\\ &= (Tf_{\tilde{X}})(0)(t(d/du)) = tX\end{aligned}$$

using $\tau^*(t)$ as a linear transformation of R for the third equality. However, the homomorphism $f_{t\tilde{X}} : R \to G$ is also such that $\dot{f}_{t\tilde{X}}(0) = tX$ so that by the uniqueness part of Theorem 5.9 we obtain $f_{t\tilde{X}}(s) = f_{\tilde{X}}(st)$ which gives the result.

Remark (1) From formula $(*)$ we see that the curve $R \to G : t \to \exp t\tilde{X}$ can be characterized as the curve $f : R \to G$ such that:

(i) f is an analytic homomorphism,
(ii) $\dot{f}(t) = \tilde{X}(f(t))$ all $t \in R$.

We also have for all $s, t \in R$ and $\tilde{X} \in \mathscr{L}(G)$ that

$$\exp(s + t)\tilde{X} = \exp s\tilde{X} \cdot \exp t\tilde{X}, \qquad [\exp t\tilde{X}]^{-1} = \exp(-t\tilde{X})$$

since $f_{\tilde{X}}$ is a homomorphism.

Theorem 5.11 Let G be a Lie group with Lie algebra $\mathscr{L}(G)$. Then the exponential map $\exp : \mathscr{L}(G) \to G$ is analytic and $T(\exp)(0) : T(\mathscr{L}(G), 0) \to T(G, e)$ is a nonsingular linear transformation.

Proof Since the vector field $\tilde{X}$ depends analytically on the parameter $X \in T(G, e)$ we have from Theorem 2.39 that there exist an open interval $(-\varepsilon, \varepsilon)$ of $0 \in R$, an open convex neighborhood U of 0 in $T(G, e)$, and an analytic mapping $u : (-\varepsilon, \varepsilon) \times U \to G$ such that for each $t \in (-\varepsilon, \varepsilon)$ and each

$X \in U$ we have $u(t, X) = f_{\tilde{X}}(t)$. Now let $a \in (-\varepsilon, \varepsilon)$ be a fixed number such that $0 < a < 1$ and let $U' = aU \subset U$. Then the map $U' \to G : X \to u(a, a^{-1}X)$ is a well-defined analytic function since $a \in (-\varepsilon, \varepsilon)$ and $a^{-1}X \in a^{-1}U' = U$. However, we have

$$u(a, a^{-1}X) = f_{a^{-1}\tilde{X}}(a) = f_{\tilde{X}}(a^{-1}a) = \exp \tilde{X}.$$

Thus using the isomorphism $\phi : \mathscr{L}(G) \to T(G, e) : \tilde{X} \to X$ of Corollary 5.3 we see $\exp \circ \phi^{-1}$ is analytic on U' since ϕ is a linear transformation which is analytic; that is, $\exp \circ \phi^{-1}$ is analytic at $0 \in T(G, e)$.

Next we shall show $\exp \circ \phi^{-1}$ is analytic on all of $T(G, e)$, for let X be any element in $T(G, e)$. Then there is a neighborhood D of X and an integer $p > 0$ so that $1/pD \subset U'$. However, since $[\exp(1/pX)]^p = \exp X$ and exp is analytic on U', we see that $\exp X$ is the (analytic) product of analytic functions so that exp is analytic on D. Finally, $T(G, e)$ is isomorphic to $\mathscr{L}(G)$ by the analytic linear transformation ϕ^{-1}. Thus $\exp = (\exp \circ \phi^{-1}) \circ \phi$ is analytic on $\mathscr{L}(G)$.

Next, to show $T(\exp)(0)$ is nonsingular we recall from Section 2.4 that the tangent space of the finite-dimensional vector space $\mathscr{L}(G)$ at the point $0 \in \mathscr{L}(G)$ equals the vector space of all directional derivatives evaluated at 0; that is, $T(\mathscr{L}(G), 0) = \{D_{\tilde{X}}(0) : \tilde{X} \varepsilon \mathscr{L}(G)\}$ and $\psi : T(\mathscr{L}(G), 0) \to \mathscr{L}(G) : D_{\tilde{X}}(0) \to \tilde{X}$ is a vector space isomorphism. Now let $f_{\tilde{X}}(t) = \exp t\tilde{X}$ and let k be any function which is analytic at $e \in G$. Then

$$\begin{aligned}[(T(\exp)(0))(D_{\tilde{X}}(0)](k) &= (D_{\tilde{X}}(0))(k \circ \exp)\\ &= \lim_{t \to 0} \frac{1}{t} [k \circ f_{\tilde{X}}(t) - k \circ f_{\tilde{X}}(0)]\\ &= \frac{d}{dt} [k(e \cdot f_{\tilde{X}}(t))]_{t=0} = (\tilde{X}k)(e) = X(k),\end{aligned}$$

where the fourth equality uses formula (*) in Section 5.1 and the fifth equality uses $\tilde{X}(e) = X$. Thus

$$T(\exp)(0) \cdot D_{\tilde{X}}(0) = X.$$

However, the isomorphisms ϕ and ψ above give $\phi\tilde{X} = X$ and $\psi(D_{\tilde{X}}(0)) = \tilde{X}$ so that $T(\exp)(0) = \phi \circ \psi$ which is nonsingular.

NOTATION In the above proof we kept track of the various vector spaces by the isomorphisms ϕ and ψ. However, to simplify notation, these vector spaces (and Lie algebras) are usually identified and the isomorphisms are ignored—it all depends on which space one takes as the definition of the Lie algebra. We shall identify as much as possible and use the Lie algebra which is most convenient. Thus we shall write $\mathscr{L}(G) = T(G, e) = g$ or equal to any other useful isomorphic characterization; for example, $\mathscr{L}(GL(V)) = T(GL(V), I) = gl(V)$. We can now consider the exponential map as defined on

$T(G, e) = g$ and note that, using $\phi = \psi = I$, Theorem 5.11 is frequently stated as follows.

Corollary 5.12 Let G be a Lie group with Lie algebra g. Then the exponential map $\exp : g \to G$ is analytic and $T(\exp)(0) = I$, the identity in $\mathrm{End}(g)$.

Example (1) For the Lie group $G = GL(V)$ we have identified its Lie algebra $\mathscr{L}(G)$ with $gl(V)$. Thus for $X \in gl(V)$ we have $f(t) = \exp tX$ as defined in this section is characterized by (i) and (ii) of remark (1) preceding Theorem 5.11. However, we have previously seen that the matrix function $e^{tX} = \sum t^n X^n/n!$ satisfies the same conditions. Thus by the uniqueness we obtain the consistent result $\exp tX = e^{tX}$.

Proposition 5.13 Let G be a Lie group with Lie algebra g. Then there exist a bounded open connected neighborhood U_0 of $0 \in g$ and an open neighborhood U_e of $e \in G$ such that $\exp : U_0 \to U_e : X \to \exp X$ is an analytic diffeomorphism.

PROOF We use the inverse function theorem as stated in Section 2.5 for analytic functions.

Definition 5.14 Let $\log : U_e \to U_0$ denote the analytic inverse of $\exp : U_0 \to U_e$ given above and let $\eta : g \to R^m$ be a vector space isomorphism. Then $D_0 = \eta(U_0)$ is open in R^m and the pair $(U_e, \eta \circ \log)$ is a chart at $e \in G$ called a **canonical** or **normal chart** at $e \in G$ and U_e is called a **canonical** or **normal neighborhood** of e.

The isomorphism η is frequently omitted and we just consider the pair $(U_e, \log)$ as a canonical chart. Thus if $X_1, \ldots, X_m$ is a basis of g, then $D_0 = \{(x_1, \ldots, x_m) \in R^m : \sum x_i X_i \in U_0\}$ and the explicit coordinate map is

$$\log : U_e \to D_0 : \exp\Big(\sum_{i=1}^{m} x_i X_i\Big) \to (x_1, \ldots, x_m).$$

As discussed for the Lie group $GL(V)$, we see in general that any point a in a Lie group G has an (analytic) chart given by $(aU_e, \log \circ L(a)^{-1})$.

REMARK (2) The image $\exp : g \to G$ is contained in the connected component of the identity G_0. However, exp need not be surjective, for let $G = SL(2, R)$ which is the Lie group of 2×2 matrices of determinant 1. We shall now briefly show there exists an element A not of the form $\exp X = e^X$. Thus let

$$A = \begin{bmatrix} r & 0 \\ 0 & 1/r \end{bmatrix}$$

where $r < -1$. Then if $A = e^X$, the characteristic roots of A are of the form e^a and e^b where a, b are characteristic roots of X. Suppose $r = e^a$ and $1/r = e^b$. Then $a = -b + 2k\pi i$. However, since $r < 0$, a is actually complex and therefore its conjugate is also a characteristic root; that is, $b = \bar{a}$. This gives a as pure imaginary. Thus we obtain a contradiction $1 = |e^a| = |r| > 1$, by the assumption $r < -1$. This contradicts $A = e^X$.

3. Exponential Formulas

In this section we use the exponential function to develop a Taylor's series expansion for a real-valued analytic function defined on a Lie group (see Section 2.8). Then using this we obtain the first few terms in the expansion of the analytic function F given by

$$\exp X \exp Y = \exp F(X, Y)$$

for X and Y in a suitable neighborhood of 0 in $T(G, e) = g$. Thus for $\theta = (0, 0)$ in $g \times g$ we obtain the local approximation for the multiplication in G by

$$F(X, Y) = X + Y + \frac{1}{2}[XY] + \frac{1}{3!}F^3(\theta)(X, Y)^{(3)} + \cdots$$

analogous to the results in Section 1.6. We shall show that the higher-order terms $F^k(\theta)(X, Y)^{(k)}$ are all contained in the subalgebra of g generated by X and Y. Finally we discuss the Campbell–Hausdorff formula for $F(X, Y)$.

Proposition 5.15 Let G be a Lie group, let f be a real-valued function analytic at $p \in G$, and let $\tilde{X} \in \mathscr{L}(G)$ with $X \in T(G, e)$. Then there exists $\varepsilon > 0$ such that, for $|t| < \varepsilon$,

$$f(p \exp tX) = \sum_{n=0}^{\infty} \frac{t^n}{n!}(\tilde{X}^n f)(p).$$

We shall refer to this formula as a **Taylor's series** expansion for f.

PROOF From formula (*) in Proposition 5.2, we have for $\alpha(t) = \exp tX$ that

$$(\tilde{X}f)(p) = [(d/dt)f(p \exp tX)]_{t=0}$$

which proves the formula

$$(\tilde{X}^n f)(p) = [(d/dt)^n f(p \exp tX)]_{t=0}$$

for $n = 1$, and we now continue by induction. Thus

$$\begin{aligned}(\tilde{X}^{n+1}f)(p) &= [\tilde{X}^n(\tilde{X}f)](p)\\ &= [(d/dt)^n(\tilde{X}f)(p \exp tX)]_{t=0}\\ &= [(d/dt)^n(d/ds)f(p \exp tX \exp sX)]_{s=0,\, t=0}\\ &= [(d/du)^n(d/du)f(p \exp uX)]_{u=0}\\ &= [(d/du)^{n+1}f(p \exp uX)]_{u=0}\end{aligned}$$

where the fourth equality uses $\exp tX \exp sX = \exp(t + s)X$ and $u = t + s$.

Now since f is analytic at p, then for some $\varepsilon > 0$ and for $|t| < \varepsilon$ we have the fact that $f(p \exp tX)$ is an analytic function of t at $0 \in R$. Thus we can write the power series

$$f(p \exp tX) = \sum_{n=0}^{\infty} a^n t^n/n!$$

for $|t| < \varepsilon$, where the a_n equal the nth derivative of $f(p \exp tX)$ at $t = 0$; that is, $a_n = (\tilde{X}^n f)(p)$, which proves the Taylor's series expansion.

Theorem 5.16 Let G be a Lie group and let X, $Y \in T(G, e)$. Then there exists $\varepsilon > 0$ such that for $|t| < \varepsilon$:

(a) $\exp tX \exp tY = \exp(tX + tY + \frac{1}{2}t^2[XY] + o(t^3))$;
(b) $\exp tX \exp tY \exp(-tX) = \exp(tY + t^2[XY] + o(t^3))$;
(c) $\exp(-tX) \exp(-tY) \exp tX \exp tY = \exp(t^2[XY] + o(t^3))$;

where in each case $o(t^3)$ is a vector in $T(G, e)$ such that for $|t| < \varepsilon$, $(1/t^3)o(t^3)$ is bounded and analytic.

PROOF (a) Let f be analytic at $p \in G$. Then using the formula

$$(\tilde{X}^n f)(p) = [(d/dt)^n f(p \exp tX)]_{t=0}$$

twice we obtain for a function g analytic at $e \in G$

$$(\tilde{X}^n \tilde{Y}^m g)(e) = [(d/dt)^n(d/ds)^m g(\exp tX \exp sY)]_{s=0,\, t=0}.$$

Therefore we obtain the Taylor's series expansion

$$g(\exp tX \exp sY) = \sum_{m,n \geq 0} \frac{t^n}{n!}\frac{s^m}{m!}[\tilde{X}^n \tilde{Y}^m g](e)$$

for s, t sufficiently near $0 \in R$. Thus for $s = t$ we obtain

$$g(\exp tX \exp tY) = \sum_{m,n \geq 0} \frac{t^{m+n}}{m!n!}[\tilde{X}^n \tilde{Y}^m g](e), \tag{1}$$

where the coefficient of t^1 is $(\tilde{X}g)(e) + (\tilde{Y}g)(e)$ and the coefficient of t^2 is $\frac{1}{2}(\tilde{X}^2 g)(e) + [\tilde{X}\tilde{Y}g](e) + \frac{1}{2}(\tilde{Y}^2 g)(e)$.

However, by Proposition 5.13 and the analyticity of the multiplication in G

$$\exp tX \exp tY = \exp F(t)$$

where $F: I \to T(G, e)$ is an analytic map of an open interval $I \subset R$ which contains $0 \in R$. Since $e = \exp 0 = \exp 0 \exp 0$, we see that $F(0) = 0$. Thus for $t \in I$, F has the Taylor's series expansion

$$F(t) = tF_1 + t^2 F_2 + o(t^3)$$

for fixed $F_1, F_2 \in T(G, e)$.

Since we are working near $e \in G$, we can assume the operations take place in a normal neighborhood U_e, and we now have for g analytic at $e \in G$ that

$$\begin{aligned} g(\exp tX \exp tY) &= g[\exp F(t)] \\ &= g[\exp(tF_1 + t^2F_2 + o(t^3))] \\ &= g[\exp(tF_1 + t^2F_2)] + o'(t^3) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} [(t\tilde{F}_1 + t^2\tilde{F}_2)^n g](e) + o'(t^3), \end{aligned} \tag{2}$$

where the last equality uses Taylor's formula for $tF_1 + t^2F_2$ and where $o'(t^3)$ denotes a real number such that for some $\varepsilon > 0$, $(1/t^3)o'(t^3)$ is analytic and bounded for $|t| < \varepsilon$. From the expression in the last equality we see that the coefficient of t^1 is $[\tilde{F}_1 g](e)$ and the coefficient of t^2 is $[\tilde{F}_2 g](e) + \frac{1}{2}[\tilde{F}_1{}^2 g](e)$.

Thus comparing the coefficients of t and t^2 in formulas (1) and (2) we obtain

$$(\tilde{F}_1 g)(e) = (\tilde{X}g)(e) + (\tilde{Y}g)(e) \qquad \text{and} \qquad (\tilde{F}_2 g)(e) = (\tfrac{1}{2}[\tilde{X}, \tilde{Y}]g)(e).$$

Since g is an arbitrary analytic function at $e \in G$ we obtain $F_1 = X + Y$ and $F_2 = \frac{1}{2}[XY]$ which shows

$$\exp tX \exp tY = \exp F(t) = \exp(tX + tY + \tfrac{1}{2}t^2[XY] + o(t^3)).$$

For t small enough, we use (a) to compute (b) as follows.

$$\begin{aligned} [\exp tX \exp tY] \exp(-tX) &= [\exp(tX + tY + \tfrac{1}{2}t^2[XY] + o(t^3))] \exp(-tX) \\ &= \exp S(t), \end{aligned}$$

where from (a)

$$\begin{aligned} S(t) &= (tX + tY + \tfrac{1}{2}t^2[XY]) + (-tX) + \tfrac{1}{2}[tX + tY + \tfrac{1}{2}t^2[XY], -tX] + o(t^3) \\ &= tY + t^2[XY] + o(t^3) \end{aligned}$$

which proves (b).

(c) Part (c) is proven similarly.

REMARKS (1) With the conditions and formulas of Theorem 5.16 we can show

(d) $\exp t(X + Y) = \exp tX \cdot \exp tY \cdot \exp o(t^2)$;
(e) $\exp(t^2[XY]) = \exp(-tX)\exp(-tY)\exp tX \exp tY \cdot \exp o(t^3)$.

Also we have the following formulas which we shall use later. For any $t \in R$,

(f) $\exp t(X + Y) = \lim_{n\to\infty}[\exp (t/n)X \exp (t/n)Y]^n$;
(g) $\exp(t^2[XY]) = \lim_{n\to\infty}\{\exp[-(t/n)X]\exp[-(t/n)Y] \exp(t/n)X \exp(t/n)Y\}^{n^2}$.

For example, to see (f), let t be fixed in R and let n be sufficiently large. Then from Theorem 5.16,

$$\exp\frac{t}{n}X\exp\frac{t}{n}Y = \exp\left[\frac{t}{n}(X + Y) + \frac{t^2}{2n^2}[XY] + o\left(\frac{1}{n^3}\right)\right]$$

and consequently

$$\left[\exp\frac{t}{n}X\exp\frac{t}{n}Y\right]^n = \exp\left[t(X + Y) + \frac{t^2}{2n}[XY] + o\left(\frac{1}{n^2}\right)\right]$$

which yields the result.

(2) Let G be a Lie group with Lie algebra g. Then the above formulas are frequently expressed as follows.

(a′) There exists a neighborhood U of 0 in g such that for all $X, Y \in U$ we have $\exp X \exp Y = \exp(X + Y + \frac{1}{2}[XY] + \varepsilon(X, Y))$ where

$$\lim_{X,Y\to 0} \varepsilon(X, Y)/\|X\|^2\,\|Y\|^2 = 0.$$

(a″) For any $X, Y \in g$ and for s, t sufficiently near $0 \in R$, $\exp sX \exp tY = \exp(sX + tY + \frac{1}{2}st[XY] + \varepsilon(s, t))$ where $\lim_{s,t\to 0}\varepsilon(s, t)/st = 0$.

(3) Now let $\mu : G \times G: (x, y) \to \mu(x, y) \equiv xy$ denote the analytic multiplication on G. Then

$$(T\mu)(e, e) : T(G, e) \times T(G, e) \to T(G, e) : (X, Y) \to X + Y,$$

for let t be near $0 \in R$. Let $\mu(\exp tX, \exp tY) = \exp F(t)$ where $F(t) = tX + tY + o(t^2)$. Then using the chain rule, $T(\exp)(0) = I$, and $d/du(0)(F) = X + Y$ we obtain the formula.

Proposition 5.17 Let G be a Lie group and let $g_1, \ldots, g_k$ be subspaces of the Lie algebra g such that g is the subspace direct sum $g_1 + \cdots + g_k$. Then the analytic map

$$\phi : g_1 + \cdots + g_k \to G : X_1 + \cdots + X_k \to (\exp X_1)\cdots(\exp X_k)$$

is such that $T\phi(0) = I$. Thus there exist bounded open connected neighborhoods of U_i of 0 in g_i and U_e of e in G such that the map

$$\phi : U_1 + \cdots + U_k \to U_e : X_1 + \cdots + X_k \to (\exp X_1)\cdots(\exp X_k)$$

is an analytic diffeomorphism.

PROOF Using induction and the inverse function theorem it suffices to show $T\phi(0) = I$ for the case $g = g_1 + g_2$. Thus let $\exp_i = \exp_G | g_i$. Then since $g_i \subset g$ we see $T(\exp_i)(0)$ is the identity map on g_i. Therefore identifying $g_1 + g_2$ with $g_1 \times g_2$ we have for $X_i \in g_i$ that $X_1 + X_2 = (X_1, X_2)$ and therefore, with μ the multiplication in G,

$$\begin{aligned}(T\phi)(0)(X_1 + X_2) &= [(T\phi)(0,0)](X_1, X_2)\\ &= [T(\mu \circ \exp_1 \times \exp_2)(0,0)](X_1 X_2)\\ &= (T\mu)(e,e)((T\exp_1)(0)X_1, (T\exp_2)(0)X_2)\\ &= X_1 + X_2,\end{aligned}$$

using $\phi = \mu \circ (\exp_1 \times \exp_2)$, the chain rule, and remark (3) above. This proves the result.

REMARKS (4) Using this result we can introduce coordinates as follows. Let $X_1, \ldots, X_m$ be a basis of g so that $X_1, \ldots, X_{r(1)}$ is a basis of g_1, $X_{r(1)+1}, \ldots, X_{r(2)}$ is a basis of g_2, etc., for $g = g_1 + \cdots + g_k$ as above. Let $U_e = \phi(U_1 + \cdots + U_k)$ be as in Proposition 5.17 so that elements of U_e are of the form

$$\phi(\sum x_i X_i) = (\exp \sum^{r(1)} x_i X_i)\cdots(\exp \sum^{r(k)} x_i X_i).$$

Let $D_0 = \{(x_1, \ldots, x_m) \in R^m : \sum x_i X_i \in U_1 + \cdots + U_k \subset g\}$. Then the map

$$U_e \to D_0 : \phi(\sum x_i X_i) \to (x_1, \ldots, x_m)$$

yields a coordinate system called a **canonical** or **normal coordinate system of the second kind**.

(5) From formula (c) of Theorem 5.16 we see that $[XY]$ is the tangent vector to the "commutator curve"

$$s \to \exp(-s^{1/2}X)\exp(-s^{1/2}Y)\exp(s^{1/2}X)\exp(s^{1/2}Y), \qquad s \geq 0$$

at $s = 0$. Thus frequently the operation $[XY]$ in g is referred to as the "commutator operation in the Lie algebra." In the work of Pontryagin [1946] this is actually used as the definition of a bilinear multiplication on $T(G, e)$ which makes it into a Lie algebra.

(6) Let U_0 be the open neighborhood of $0 \in g$ as given in Theorem 5.13.

Then since $\exp : U_0 \to U_e$ is a diffeomorphism we can define a composition $F : U_0 \times U_0 \to g$ by

$$F(X, Y) = \log(\exp X \exp Y)$$

for $X, Y \in U_0$ provided $\exp X \exp Y \in U_e$. Thus U_0 with this composition becomes a local Lie group with 0 as the identity and $-X$ as the inverse of X. Now for any $X, Y \in g$ and t sufficiently near $0 \in R$ we have from formula (a),

$$F(tX, tY) = tX + tY + \tfrac{1}{2}t^2[XY] + o(t^3),$$

where the error term $o(t^3)$ is also a function of $X, Y \in g$. Actually, $o(t^3)$ can be expressed quite nicely by a series which is contained in the Lie algebra generated by X and Y; that is, looking ahead to Chapter 6, we define a Lie subalgebra h of the Lie algebra g to be a subvector space of g such that for all $X, Y \in h$ we have $[XY] \in h$. In particular if $L(X, Y)$ denotes the subalgebra of g generated by X and Y; that is,

$$L(X, Y) = \bigcap \{h : h \text{ is a Lie subalgebra of } g \text{ and } X, Y \in h\},$$

then $L(X, Y)$ is spanned as a vector space by X, Y, $[XY]$, $[X[XY]]$, $[Y[YX]]$, etc. Thus we shall show that the above formula has the form

$$F(tX, tY) = \sum_{i=1}^{\infty} F_i(tX, tY)$$

where each $F_i(tX, tY) \in L(X, Y)$. This result is known as the "Campbell–Hausdorff theorem." Note that since a subvector space is closed we also have $F(tX, tY) \in L(X, Y)$. Consequently the multiplication in the neighborhood of U_e is completely determined by the Lie algebra. We shall express this more accurately later by noting that two Lie groups are locally isomorphic if their corresponding Lie algebras are isomorphic.

Exercises (1) As in the proof of Theorem 5.16 we can write $\exp tX \cdot \exp tY = \exp F(t)$, where

$$F(t) = tF_1 + t^2F_2 + t^3F_3 + o(t^4).$$

Show that the third-order term is

$$F_3 = \tfrac{1}{12}\{[X[XY]] + [Y[YX]]\}.$$

(2) Let G be a Lie group with Lie algebra $g = T(G, e)$.

(i) If $X, Y \in g$ are such that $[XY] = 0$, then show $\exp X \exp Y = \exp(X + Y)$. A Lie algebra is called **commutative** or **Abelian** if $[XY] = 0$ for all $X, Y \in g$.

(ii) Let G be a connected Lie group with Lie algebra g. Then G is commutative if and only if g is commutative.

We shall now prove that Campbell–Hausdorff theorem as mentioned in remark (6); this proof is essentially that given by Eichler [1968]. Following his notation we shall call an expression $G(X, Y, \dots, Z)$ a **Lie polynomial** if $G(X, Y, \dots, Z) \in L(X, Y, \dots, Z)$ which is the Lie subalgebra of g generated by $X, Y, \dots, Z$ in g. Thus $G(X, Y)$ is a Lie polynomial if it is a linear combination of X, Y, $[XY]$, etc. An expression $G(X, Y, \dots, Z)$ is **homogeneous of degree** k if for any $s \in R$, $G(sX, sY, \dots, sZ) = s^k G(X, Y, \dots, Z)$. By previous results of this section we have for $X, Y \in g$ and s near enough $0 \in R$ that

$$\exp sX \cdot \exp sY = \exp F(sX, sY),$$

where F is analytic at $\theta = (0, 0) \in g \times g$ and has the series expansion

$$F(sX, sY) = \sum_{k=1}^{\infty} \frac{s^k F^k(\theta)}{k!} (X, Y)^{(k)}.$$

Thus $F^1(\theta)(X, Y) = X + Y$ and $F^2(\theta)(X, Y)^{(2)} = [XY]$ are Lie polynomials which are homogeneous of degree 1 and 2, respectively. We can now state the result as follows.

Theorem 5.18 (Campbell–Hausdorff theorem) Let G be a Lie group with Lie algebra g. Let $X, Y \in g$ and let s be sufficiently near 0 in R so that the multiplication is represented locally by

$$\exp sX \exp sY = \exp F(sX, sY)$$

where the analytic function F is given by the above series. Then each $F^k(\theta)(X, Y)^{(k)}$ is a Lie polynomial in X and Y and is homogeneous of degree k.

PROOF We first note the following facts which we leave as brief exercises.

(a) For $X, Y \in g$ and s near enough $0 \in R$,

$$F^k(\theta)(sX + o(s^2), sY)^{(k)} = s^k F^k(\theta)(X, Y)^{(k)} + o(s^{k+1}).$$

[Just use the multilinearity of $F^k(\theta)$ and write $(sX + o(s^2), sY) = (sX, sY) + (o(s^2), 0)$.]

(b) If $G(X, Y, \dots, Z), H(X, Y, \dots, Z), \dots, K(X, Y, \dots, Z) \in L(X, Y, \dots, Z)$; that is, Lie polynomials in $X, Y, \dots, Z$, then $G(H(X, Y, \dots, Z), \dots, K(X, Y, \dots Z))$ is also a Lie polynomial in $X, Y, \dots, Z$.

(c) If $G(X, Y, \dots, Z)$ is a Lie polynomial and $G(X, Y, \dots, Z) = \sum_k G_k(X, Y, \dots, Z)$, where the $G_k(X, Y, \dots, Z)$ are the homogeneous components of degree k into which G decomposes, then each $G_k(X, Y, \dots, Z)$ is a Lie polynomial.

We shall now use the notation

$$F_k(sX, sY) = \frac{s^k}{k!} F^k(\theta)(X, Y)^{(k)}$$

and show by induction that each $F_k(sX, sY)$ is a Lie polynomial in X and Y. For $k = 1, 2$ this has already been done and we assume for all $k < n$ that each $F_k(sX, sY)$ is a Lie polynomial in X and Y. From the associative law

$$(\exp sX \exp sY) \exp sZ = \exp sX(\exp sY \exp sZ)$$

we obtain

$$\sum_{i=1}^{\infty} F_i\Big(\sum_{j=1}^{\infty} F_j(sX, sY), sZ\Big) = \sum_{i=1}^{\infty} F_i\Big(sX, \sum_{j=1}^{\infty} F_j(sY, sZ)\Big). \tag{1}$$

Expanding the left side of (1) we see [using (a) and $F_1(X, Y) = X + Y$] that the homogeneous term of degree n is

$$s^n\{F_n(X, Y) + \tfrac{1}{2}[F_{n-1}(X, Y)Z] + \cdots + F_n(X + Y, Z)\}.$$

Thus by the induction hypothesis, the only possible term in this expression which might not be a Lie polynomial is

$$s^n\{F_n(X, Y) + F_n(X + Y, Z)\}.$$

Similarly expanding the right side of (1) we see that the only possible homogeneous term of degree n which might not be a Lie polynomial is

$$s^n\{F_n(X, Y + Z) + F_n(Y, Z)\}.$$

This relation can be expressed as follows: We write $U \sim V$ if $U - V$ is a Lie polynomial. Thus we have

$$F_n(X, Y) + F_n(X + Y, Z) \sim F_n(X, Y + Z) + F_n(Y, Z) \tag{2}$$

and using (2) we will show

$$F_n(X, Y) \sim 0;$$

that is, $F_n(X, Y)$ is a Lie polynomial in X and Y.

We now compute. First from $\exp aX \exp bX = \exp(a + b)X$ we see $F(asX, bsX) = (a + b)sX$ so that for $k > 1$,

$$F_k(aX, bX) = 0 \tag{3}$$

and using $\exp X \exp 0 = \exp X$ we also have for $k > 1$

$$F_k(X, 0) = F_k(0, X) = 0. \tag{4}$$

We now omit the s near $0 \in R$ for the following computations and let $Z = -Y$ in (2) to obtain

$$F_n(X, Y) + F_n(X+Y, -Y) \sim F_n(X, Y-Y) + F_n(Y, -Y)$$

so that by (3) and (4),

$$F_n(X, Y) \sim -F_n(X+Y, -Y). \tag{5}$$

Let $X = -Y$ in (2) to obtain $0 \sim F_n(-Y, Y+Z) + F_n(Y, Z)$. Thus replacing Y by X and Z by Y we obtain

$$F_n(X, Y) \sim -F_n(-X, X+Y). \tag{6}$$

This yields

$$\begin{aligned} F_n(X, Y) &\sim -F_n(-X, X+Y), && \text{using (6)} \\ &\sim -(F_n(-X+(X+Y)), -(X+Y)), && \text{using (5)} \\ &= F_n(-(-Y), -(X+Y)) \\ &= (-1)^n F_n(-Y, X+Y), && \text{homogeneity} \\ &\sim -(-1)^n F_n(Y, X), && \text{using (6).} \end{aligned} \tag{7}$$

Next Let $Z = -\frac{1}{2}Y$ in (2) to obtain

$$F_n(X, Y) \sim -F_n(X+Y, -\tfrac{1}{2}Y) + F_n(X, \tfrac{1}{2}Y). \tag{8}$$

Similarly let $X = -\frac{1}{2}Y$ in (2) to obtain $F_n(\frac{1}{2}Y, Z) \sim F_n(-\frac{1}{2}Y, Y+Z) + F_n(Y, Z)$ so that replacing Y by X and Z by Y we obtain

$$F_n(X, Y) \sim F_n(\tfrac{1}{2}X, Y) - F_n(-\tfrac{1}{2}X, X+Y). \tag{9}$$

We now use (8) on the two terms on the right of (9) to obtain

$$\begin{aligned} F_n(X, Y) \sim &-F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}Y) + F_n(\tfrac{1}{2}X, \tfrac{1}{2}Y) + F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}X-\tfrac{1}{2}Y) \\ &- F_n(-\tfrac{1}{2}X, \tfrac{1}{2}X+\tfrac{1}{2}Y) \\ = &(\tfrac{1}{2})^n F_n(X, Y) - (\tfrac{1}{2})^n F_n(-X, X+Y) - F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}Y) \\ &+ F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}X-\tfrac{1}{2}Y). \end{aligned} \tag{10}$$

Next

$$\begin{aligned} F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}Y) &\sim -F_n(\tfrac{1}{2}X+Y+(-\tfrac{1}{2}Y), \tfrac{1}{2}Y), && \text{using (5)} \\ &= -F_n(\tfrac{1}{2}(X+Y), \tfrac{1}{2}Y) \\ &= -(\tfrac{1}{2})^n F_n(X+Y, Y) \end{aligned} \tag{11}$$

and

$$\begin{aligned} F_n(\tfrac{1}{2}X+Y, -\tfrac{1}{2}X-\tfrac{1}{2}Y) &\sim -F_n(\tfrac{1}{2}X+Y+(-\tfrac{1}{2}X-\tfrac{1}{2}Y), \tfrac{1}{2}X+\tfrac{1}{2}Y) \\ &= -(\tfrac{1}{2})^n F_n(Y, X+Y) \end{aligned} \tag{12}$$

where we again use (5). Thus, since from (6) we have $-F_n(-X, X+Y) \sim F_n(X, Y)$, we substitute (11) and (12) into (10) to obtain

$$F_n(X, Y) \sim (\tfrac{1}{2})^{n-1}F_n(X, Y) + (\tfrac{1}{2})^n F_n(X+Y, Y) - (\tfrac{1}{2})^n F_n(Y, X+Y). \quad (13)$$

Thus, using (7) we have from (13)

$$[1 - (\tfrac{1}{2})^{n-1}]F_n(X, Y) \sim (\tfrac{1}{2})^n(1 + (-1)^n)F_n(X+Y, Y). \quad (14)$$

In case $n > 1$ is odd we obtain the desired result $F_n(X, Y) \sim 0$. In case n is even we replace X by $X - Y$ in (5) and use (14) to obtain

$$\begin{aligned} -F_n(X, -Y) &\sim F_n(X - Y, Y) \\ &\sim (\tfrac{1}{2})^n(1 + (-1)^n[1 - (\tfrac{1}{2})^{n-1}]^{-1}F_n(X, Y). \end{aligned} \quad (15)$$

We apply (15) twice to obtain

$$\begin{aligned} F_n(X, Y) &= -(-F_n(X, -(-Y))) \\ &\sim -(\tfrac{1}{2})^n(1 + (-1))^n[1 - (\tfrac{1}{2})^{n-1}]^{-1}F_n(X, -Y) \\ &\sim (\tfrac{1}{2})^{2n}(1 + (-1)^n)^2[1 - (\tfrac{1}{2})^{n-1}]^{-2}F_n(X, Y). \end{aligned}$$

Since the coefficient on the right in (16) equals $1/(2^{n-1} - 1)^2 \neq 1$, we again obtain $F_n(X, Y) \sim 0$. This completes the proof of the Campbell–Hausdorff theorem.

REMARKS (6) The above proof is a relatively straightforward computation not involving much machinery. However, with more development the proofs in Jacobson [1962] and Serre [1965] give the following explicit formula (Campbell–Hausdorff formula): For a suitable neighborhood U_0 of 0 in g we have for $X, Y \in U_0$ that

$$F(X, Y) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\left[\sum \frac{\tau(X^{p_1}, Y^{q_1}, X^{p_2}, Y^{q_2}, \ldots, X^{p_n}, Y^{q_n})}{C(p_1, q_1, \ldots, p_n, q_n)}\right],$$

where the second sum runs over the integers $p_i, q_i \geq 0$ with $p_i + q_i \geq 1$ for $i = 1, \ldots, n$ and

$$C(p_1, q_1, \ldots, p_n, q_n) = \sum_{i=1}^{n}(p_i + q_i)p_1!q_1!\cdots p_n!q_n!,$$

$$\tau(X^{p_1}, Y^{q_1}, \ldots, X^{p_n}, Y^{q_n}) = [[\cdots[[[X^{p_1}Y^{q_1}]X^{p_2}]Y^{q_2}]\cdots X^{p_n}]Y^{q_n}]$$

where we use the notation

$$[U^rV^s] = [[\cdots[[[\underbrace{UU]U]\cdots U}_{r}]\underbrace{V]\cdots V}_{s}]$$

for $k > 1$ and $\tau(X^1) = X$, $\tau(Y^1) = Y$.

If we let $\operatorname{ad}(X) : g \to g : Y \to [XY]$, then the above expression for τ is

$$\tau(X^{p_1}, Y^{q_1}, \ldots, X^{p_n}, Y^{q_n}) = (\operatorname{ad} X)^{p_1}(\operatorname{ad} Y)^{q_1} \cdots (\operatorname{ad} Y)^{q_n - 1} Y$$

if $q_n \geq 1$ and

$$\tau(X^{p_1}, Y^{q_1}, \ldots, X^{p_n}) = (\operatorname{ad} X)^{p_1}(\operatorname{ad} Y)^{q_1} \cdots (\operatorname{ad} X)^{p_n - 1} X$$

if $q_n = 0$. From these formulas observe that if $q_n \geq 2$, or if $q_n = 0$ and $p_n \geq 2$, then $\tau(X^{p_1}, \ldots, X^{p_n}, Y^{q_n})$ is 0. Thus possible nonzero terms occur only when $q_n = 1$ or when $q_n = 0$ and $p_n = 1$.

Exercise (3) Compute the term $F^4(\theta)(X, Y)^{(4)}/4!$ using the Campbell–Hausdorff formula.

4. Homomorphisms and Analytic Structure

We use previous results of this chapter to discuss elementary facts on homomorphisms from which we obtain the fact that the topology of a Lie group uniquely determines its analytic structure.

Proposition 5.19 Let G and H be Lie groups with Lie algebras g and h, respectively, and let $f: G \to H$ be a group homomorphism which is an analytic map of manifolds; that is, an analytic homomorphism. Then $Tf(e) : g \to h$ is a Lie algebra homomorphism and for $X \in g$

$$f(\exp_G X) = \exp_H[(Tf(e))(X)].$$

Thus for $\bar{e}$ the identity in H, the accompanying diagram is commutative.

$$\begin{array}{ccc} T(G, e) & \xrightarrow{(Tf)(e)} & T(H, \bar{e}) \\ \Big\downarrow{\scriptstyle \exp_G} & & \Big\downarrow{\scriptstyle \exp_H} \\ G & \xrightarrow[f]{} & H \end{array}$$

PROOF For $X \in g$ the map

$$\phi : R \to H : t \to f(\exp tX)$$

is an analytic homomorphism and for $\alpha(t) = \exp tX$ we see

$$\begin{aligned} \dot{\phi}(0) &= [T(f \circ \alpha)(d/du)](0) \\ &= (Tf)(e)\dot{\alpha}(0) = (Tf)(e)X. \end{aligned}$$

However $\beta : R \to H : t \to \exp_H t(Tf(e)X)$ is an analytic homomorphism and $\dot{\beta}(0) = Tf(e)X$. Thus by the uniqueness of Theorem 5.9 we obtain $f(\exp tX) = \exp t(Tf(e)X)$. Next, since $Tf(e) : g \to h$ is a vector space homomorphism, we must show $Tf(e)[X\ Y] = [Tf(e)X\ Tf(e)Y]$ for all X, $Y \in g$. Notice that the two maps $p : G \times G \to H : (x, y) \to f(x^{-1}y^{-1}xy)$ and $g : G \times G \to H : (x, y) \to f(x)^{-1}f(y)^{-1}f(x)f(y)$ are equal. Thus for t near $0 \in R$ we apply this to $x = \exp tX$, $y = \exp tY$ and use Theorem 5.16(c) together with the formula of the preceding paragraph to obtain $Tf(e)$ as a Lie algebra homomorphism.

Exercises (1) Let $f: G \to H$ be an analytic homomorphism of Lie groups. Then $Tf(e) : g \to h$ is surjective (injective) if and only if $Tf(a) : T(G, a) \to T(H, f(a))$ is surjective (injective) for all $a \in G$. Thus f is an immersion or submersion depending only upon the value $Tf(p)$ at the single point $p = e$.

(2) If in (1) the map $f: G \to H$ is injective, then f is regular; that is, $Tf(a)$ is injective for all $a \in G$. (Note Proposition 5.19.)

(3) If G is a commutative connected Lie group with Lie algebra g, then $\exp : g \to G$ is surjective. (See exercises in Section 5.3 and regard exp as a homomorphism of a commutative groups.)

REMARK (1) From the preceding results we see that an analytic isomorphism of Lie groups implies an isomorphism of the corresponding Lie algebras. The converse is true locally, and globally if the groups are simply connected. Consider the locally isomorphic groups R and T^1 for a counterexample. It is this converse which allows a classification of certain Lie groups by classifying their Lie algebras.

Proposition 5.20 Let α, β be two analytic homomorphisms of the Lie group G into the Lie group H. If $T\alpha(e) = T\beta(e)$, then there exists an open neighborhood U of e in G on which α and β are equal. Furthermore, if G is connected, then $\alpha = \beta$ on G.

PROOF Let $(U_e, \log)$ be a canonical chart at $e \in G$ and set $U = U_e$. Then from Proposition 5.19 we have for $a = \exp X \in U$, that

$$\begin{aligned}\alpha(a) &= \alpha(\exp(X))\\ &= \exp[(T\alpha(e))(X)]\\ &= \exp[(T\beta(e))(X)] = \beta(a)\end{aligned}$$

which proves part of the result. The other part follows from the fact that U generates G and α, β are homomorphisms.

The first fact needed to prove that the topology of a Lie group determines is analytic structure is the following result.

Lemma 5.21 Let G be a Lie group with Lie algebra g, let $f: R \to G$ be a homomorphism of the additive group R into the group G, and let f be a continuous map of the corresponding topological spaces. Then these exists $X \in g$ such that $f(t) = \exp tX$. Thus a continuous homomorphism $f: R \to G$ is actually analytic.

PROOF Let $(U, \log)$ be a canonical chart at $e \in G$ and let V be a neighborhood of e in G with $VV \subset U$. We shall now show that the map $V \to U$: $a \to a^2$ is injective. For if $a \in V$, then $a^2 \in U$ so that $\log a$ and $\log a^2$ are defined. Thus $\phi : R \to G : t \to \exp(t \log a)$ is a homomorphism such that $a^2 = \exp(2 \log a)$. However, since $a^2 \in U$, $a^2 = \exp(\log a^2)$ and since $\exp | U$ is injective, $2 \log a = \log a^2$. Thus $a = \exp(\log a) = \exp(\frac{1}{2} \log a^2)$ so that $a^2 \in U$ uniquely determines $a \in V$. Thus "square roots" exist and are unique.

Next since the given homomorphism $f: R \to G$ is continuous, there exists $\varepsilon > 0$ such that $f(t) \in V$ if $|t| \leq \varepsilon$. Now we can assume $\varepsilon = 1$ otherwise we can make a change of parameter to $s = \lambda t$ where now $f(s)$ is defined for $|s| \leq 1$. Using this let

$$a = f(1) \in V \qquad \text{and} \qquad X = \log a \in g.$$

From the preceding paragraph $\exp \frac{1}{2}X$ is the unique square root of $a \in V$. Thus since $f(\frac{1}{2})$ is also a square root of $a = f(1)$ we obtain by uniqueness $f(\frac{1}{2}) = \exp \frac{1}{2}X$. Applying this argument to $f(\frac{1}{2})$ and taking its unique square root we obtain $f(\frac{1}{4}) = \exp \frac{1}{4}X$ and therefore $\log f(\frac{1}{4}) = \frac{1}{4}X$. Using induction one obtains $f(1/2^n) = \exp(1/2^n X)$ and for any integer p we have since f is a homomorphism

$$f(p/2^n) = f(1/2^n)^p = \exp((1/2^n)X)^p = \exp((p/2^n)X).$$

Thus for every dyadic rational number q we have $f(q) = \exp qX$ and by continuity $f(t) = \exp tX$ for all $t \in R$.

Analogous to Corollary 3.15 we have the following result.

Exercise (4) Let $f: G \to H$ be an (algebraic) homomorphism of the Lie groups G and H. If f is analytic at $e \in G$, then f is analytic on all of G.

Theorem 5.22 Let G, H be Lie groups and let $f: G \to H$ be a continuous homomorphism of the corresponding topological groups. Then f is an analytic homomorphism of the Lie groups.

PROOF Let g be the Lie algebra of G and let $g = g_1 + \cdots + g_m$ be a subspace direct sum of the one-dimensional spaces $g_i = RX_i$ where $X_1, \ldots, X_m$ is some basis of g. Let $\exp_i = \exp | g_i$ and let

$$\phi : g \to G : A_1 + \cdots + A_m \to \exp A_1 \cdots \exp A_m$$

be the map which gives the canonical coordinates of the second kind where $A_i = t_i X_i \in g_i$ [see remark (4), Section 5.3]. Then, since

$$f \circ \phi(A_1 + \cdots + A_m) = f(\exp_1 A_1 \cdots \exp_m A_m) = f(\exp_1 A_1) \cdots f(\exp_m A_m),$$

we have an expression for $f \circ \phi : g \to H$ in terms of analytic functions, applying Lemma 5.21 to each $f \circ \exp_i$ for $i = 1, \ldots, m$. Thus $f \circ \phi : g \to H$ is analytic. However ϕ is a local analytic diffeomorphism of a neighborhood U_0 of 0 in g into a neighborhood U_e of e in G. Therefore, since $\phi^{-1} : U_e \to U_0$ is analytic, $f = (f \circ \phi) \circ \phi^{-1}$ is analytic on U_e; that is, f is analytic at $e \in G$, and by the preceding exercise f is analytic on all of G.

Corollary 5.23 Two real Lie groups which are isomorphic as topological groups are actually isomorphic as Lie groups.

REMARK (2) Let G be a Lie group with analytic structures $\mathscr{A}_1$ and $\mathscr{A}_2$ which give topologies $\mathscr{T}_1$ and $\mathscr{T}_2$. Let $\mathscr{G}_i$ denote the Lie groups $(G, \mathscr{A}_i)$ and let G_i denote the topological groups $(G, \mathscr{T}_i)$. Then $G_1 \cong G_2$ as topological groups implies $\mathscr{G}_1 \cong \mathscr{G}_2$ as Lie groups; that is, the continuous isomorphism which expresses $G_1 \cong G_2$ is actually analytic and therefore an analytic isomorphism which expresses $\mathscr{G}_1 \cong \mathscr{G}_2$. Thus the analytic structures are uniquely determined. In particular, if we say $G_1 = G_2$ via the continuous identity isomorphism, then $\mathscr{G}_1 = \mathscr{G}_2$ as Lie groups.

CHAPTER 6

LIE SUBGROUPS AND SUBALGEBRAS

In this chapter we consider Lie subalgebras of a Lie algebra and show their basic relationships with Lie subgroups. Thus each Lie subgroup yields a Lie subalgebra and conversely each Lie subalgebra is the Lie algebra of a Lie subgroup. Next we show that two Lie groups are locally isomorphic if and only if their Lie algebras are isomorphic; in Chapter 8 we extend this to a global result. In the third section we prove the very useful result that an abstract subgroup of a Lie group, which is a closed subset, is actually a Lie subgroup. Next we discuss homogeneous spaces G/H where H is a closed subgroup of the Lie group G and show how to coordinatize G/H using the exponential mapping, Then we apply these results to quotient groups. Finally we show that a commutative connected Lie group is isomorphic to $R^q \times T^p$ where R^q is a q-dimensional Euclidean space and T^p is a p-dimensional torus.

1. Lie Subalgebra and Uniqueness of Analytic Structure

From the various characterizations of a Lie subgroup H of a Lie group G, we see that the Lie subgroup $H \subset G$ must satisfy:

(1) H is a Lie group;
(2) the injection $f: H \to G : x \to x$ is an analytic immersion.

Thus $(Tf)(b) : T(H, b) \to T(G, b)$ is injective for every $b \in H$. We use $Tf(e)$ to identify the Lie algebra of H with a subalgebra of the Lie algebra of G and show how a Lie subalgebra can be used to generate a Lie subgroup of G.

Definition 6.1 Let g be an (abstract) Lie algebra; that is, g has an anticommutative bilinear multiplication $[XY]$ such that the Jacobi identity holds (Section 1.6). Then a subvector space $h \subset g$ is a **Lie subalgebra** if for all $X, Y \in h$ we have $[XY] \in h$.

Proposition 6.2 Let H be a Lie subgroup of the Lie group G given by $f: H \to G$ as above. Then $Tf(e) : T(H, e) \to T(G, e)$ is an injective homomorphism. Thus we can consider the Lie algebra of H as a Lie subalgebra of the Lie algebra of G.

PROOF We just note that the identity map $f: H \to G : x \to x$ is a homomorphism and therefore $Tf(e)$ is a Lie algebra homomorphism. Also, by definition of immersion. $Tf(e)$ is injective.

Corollary 6.3 Let H be a Lie subgroup of a Lie group G and let h, g be the corresponding Lie algebras and regard $h \subset g$. Then $\exp_H = \exp_G | h$.

PROOF Let $f: H \to G : x \to x$, Then from Proposition 5.19 we have for $X \in T(H, e)$,

$$\exp_H X = f(\exp_H X) = \exp_G\big(Tf(e)X\big).$$

Thus after the above identifications, we obtain the result.

Proposition 6.4 Let G be a Lie group with Lie algebra g and let H be a Lie subgroup of G with Lie algebra h. Then $h = \{X \in g : \text{all } t \in R, \exp tX \in H$ and $R \to G : t \to \exp tX$ is continuous$\}$.

PROOF For $X \in h$, we have from Lemma 5.21 that $\alpha : R \to H : t \to \exp_H tX = \exp_G tX$ is analytic which gives one inclusion. Conversely, suppose for $X \in G$ that the map $\alpha : R \to H : t \to \exp_G tX$ is continuous. Then from Lemma 4.9 (with $f = \exp_G$, $M = R$, $N = G$, and $P = H$), we see that α is analytic. Thus since $\alpha(t) \in H$ with $\alpha(0) = e \in H$, we obtain $X = \dot{\alpha}(0) \in T(H, e) = h$.

Exercise (1) Let H_1 and H_2 be connected Lie subgroups of a Lie group G, and let h_1 and h_2 be the corresponding Lie subalgebras. If $h_1 = h_2$, then $H_1 = H_2$ as Lie groups.

From the above results or from Corollary 5.23, we have the following:

Corollary 6.5 If H_1 and H_2 are two Lie subgroups of the Lie group G such that H_1 and H_2 are equal as topological groups. then they are equal as Lie groups.

Theorem 6.6 Let G be a Lie group with Lie algebra g. If H is a Lie subgroup of G with Lie algebra h, then h is a Lie subalgebra of g. Conversely, for each Lie subalgebra h of g, there exists a unique connected Lie subgroup H of G which has h as its Lie algebra. This Lie subgroup H is the smallest Lie subgroup of G containing $\exp h$.

PROOF From the previous results, it suffices to prove the converse. First, the uniqueness follows from the above exercise. We now use the Campbell–Hausdorff theorem to obtain a local Lie group B of G as follows. Let m be a subspace of g such that $g = m + h$ (subspace direct sum). Then using canonical coordinates of the second kind [Section 5.3, remark (4)], there exist open, connected, symmetric neighborhoods V (respectively W) of 0 in h (respectively m) and a neighborhood U of e in G such that the map

$$\phi : W + V \to U : X + Y \to \exp X \cdot \exp Y$$

is an analytic diffeomorphism. Now let $B = \exp V$ where $V \subset h$. Then since h is a subalgebra, we have for all Y and Z in a suitable neighborhood of 0 in h, $F(Y, Z) = \log(\exp Y \exp Z) \in h$ which uses the Campbell–Hausdorff theorem. Thus using Theorem 5.16 we see that B is a local Lie group relative to the operations in G (brief exercise). Note that the topology on B is given by prescribing as a family of neighborhoods of $e \in B$ the system of sets $\exp S$ where S ranges over the neighborhoods of 0 in V.

Next, using the canonical coordinates of the second kind as given above, we see (U, ϕ^{-1}) is a chart at e in G. Also defining $y : B \to V : \exp Y \to Y$ for $Y \in h$, we obtain a chart (B, y) of e in B such that $\phi^{-1} \mid B = y$ and the Jacobian matrix $\left(\partial(\phi^{-1})_i/\partial y_j(e)\right)$ has rank equal to $\dim B$ ($= \dim h$). Thus by Corollary 4.13, B generates a connected Lie subgroup H of G.

Let $\mathscr{L}(H)$ denote the Lie algebra of H. Then from Proposition 6,4 we see $h \subset \mathscr{L}(H)$. However, since $\dim h = \dim B = \dim H = \dim \mathscr{L}(H)$, we have equality.

REMARK (1) We can use Theorem 6.6 to show that an (abstract) finite-dimensional Lie algebra L over R occurs (up to an isomorphism) as the Lie algebra of some Lie group. First we state Ado's theorem (see Jacobson's proof [1962]).

Ado's Theorem Let L be a finite-dimensional Lie algebra over a field K. Then there exists some finite-dimensional vector space V over K such that L is isomorphic to a Lie subalgebra h of $gl(V)$.

Now $gl(V)$ is the Lie algebra of $GL(V)$, so that the connected subgroup H of $GL(V)$ generated by $\exp h$ is a Lie group with Lie algebra h isomorphic to L. Note that this does not give much information on L but just says that $gl(V)$ and $GL(V)$ are rather complicated.

2. Local Isomorphisms

We now determine what type of information about a Lie group can be obtained from its Lie algebra—roughly those properties preserved by a Lie algebra isomorphism.

Proposition 6.7 Let $f: G \to H$ be a local homomorphism of the Lie groups G and H. Let g and h be the corresponding Lie algebras and assume the Lie algebra homomorphism $(Tf)(e) : g \to h$ is bijective. Then f is a local isomorphism.

PROOF Since $Tf(e)$ is bijective, we have by the inverse function theorem neighborhoods U of e in G and V of $\bar{e}$ in H such that $f: U \to V$ is an analytic diffeomorphism. However, since f is a local homomorphism, we have on suitable neighborhoods that the analytic inverse of f is also a local homomorphism; that is, f is a local isomorphism.

Theorem 6.8 Let G and H be Lie groups with Lie algebras g and h.

(a) If $f: g \to h$ is a Lie algebra homomorphism, then there exists a local Lie group homomorphism $\phi : G \to H$ such that $(T\phi)(e) = f$.

(b) G and H are locally isomorphic as Lie groups if and only if g and h are isomorphic as Lie algebras.

PROOF Part (b) follows from (a), Proposition 6.7, and the obvious modification of Proposition 5.19. For (a) we follow the proof of Chevalley [1946, p. 112] and note that $k = \{(X, f(X)) : X \in g\}$ is a Lie subalgebra of $g \times h$ as in exercise (2) in Section 5.1. Let K be the connected subgroup of $G \times H$ with Lie algebra k (Theorem 6.6), and let $\pi : G \times H \to G : (x, y) \to x$ be the analytic projection map. Now the map $r = \pi | K : K \to G$ is a Lie group homomorphism and

$$(Tr)(e, \bar{e}) : k \to g : (X, f(X)) \to X$$

is actually a Lie algebra isomorphism since $X = 0$ implies $(X, f(X)) = (0, 0)$ and the map is clearly surjective. Thus by Proposition 6.7, r is a local isomorphism with local inverse $s : V \to W$ where $V \subset G$ and $W \subset K$ are suitable neighborhoods of the corresponding identity elements. Also we have $(Ts)(e) : g \to k$ and

$$\begin{aligned} Ts(e)X &= (Ts)(e)[(Tr)(e, \bar{e})(X, f(X))] \\ &= T(s \circ r)(e, \bar{e})(X, f(X)) = (X, f(X)). \end{aligned}$$

Now let $p : G \times H \to H$ be the natural analytic projection. Then p is a Lie group homomorphism. Thus if we let

$$\phi = p \circ s : G \to H,$$

we obtain the desired local homomorphism because

$$\begin{aligned}(T\phi)(e)X &= T(p \circ s)(e)X \\ &= (Tp)(e, \bar{e})\big(X, f(X)\big) = f(X).\end{aligned}$$

REMARKS (1) As previously stated, we shall use simple connectivity to obtain the following result: Let G be a connected and simply connected Lie group, and let $f: G \to H$ be a local homomorphism of G into a Lie group H. Then there exists a unique extension of f to a (global) homomorphism of G into H.

(2) The proof of Theorem 6.8 uses indirectly the Campbell–Hausdorff theorem, and we now sketch a more direct approach. Using the notation of Section 5.3, let U_0 be a neighborhood of $0 \in g$ and U_e a neighborhood of $e \in G$ such that we have the composition for X, $Y \in U_0$,

$$\exp X \exp Y = \exp F(X, Y),$$

where $F(X, Y) = X + Y + \frac{1}{2}[XY] + \cdots + 1/k!\ F^k(\theta)(X, Y)^{(k)} + \cdots$ is given by the Campbell–Hausdorff theorem. From this theorem each term $F^k(\theta)(X, Y)^{(k)}$ is a Lie polynomial in X and Y. Therefore, for the homomorphism $f: g \to h$ we have $f[F^k(\theta)(X, Y)^{(k)}] = F^k(\theta)\big(f(X), f(Y)\big)^{(k)}$, using the definition of a Lie polynomial. Because f is linear (and continuous) this gives

$$f[F(X, Y)] = F\big(f(X), f(Y)\big).$$

We can assume that U_0 is small enough in g so similar results hold in some neighborhood of $0 \in h$ which contains $f(U_0)$. Now for $x = \exp X \in U_e$, define ϕ by

$$\phi(x) = \exp_H f(X)$$

and note that for $x = \exp X$, $y = \exp Y \in U_e$

$$\begin{aligned}\phi(xy) &= \phi\big(\exp F(X, Y)\big) \\ &= \exp_H f[F(X, Y)] \\ &= \exp_H F\big(f(X), f(Y)\big) \\ &= \exp_H f(X) \cdot \exp_H f(Y) = \phi(x) \cdot \phi(y).\end{aligned}$$

Also we see $T\phi(e)X = f(X)$.

3. Closed Subgroups

We have previously given criterion for a subgroup H of a Lie group G to be a Lie subgroup, and we now give an extremely useful criterion concerning a closed subgroup.

Theorem 6.9 Let G be a Lie group, and let H be an (abstract) subgroup of G which is a closed subset of the topological space G. Then there exists a unique analytic structure on H such that H is a Lie subgroup of G and the topology on H induced by this analytic structure is the topology induced by G; that is, H is a topological subgroup of G. Furthermore the Lie algebra of H equals $\{X \in g : \exp_G tX \in H \text{ for all } t \in R\}$.

PROOF First note that the uniqueness of this analytic structure follows from Corollary 6.5. From the definition of a Lie subgroup, it suffices to show H is an analytic submanifold of G. We shall use the various group properties and closure, and separate the proof into several parts.

(a) Let $h = \{X \in g : \exp tX \in H \text{ for all } t \in R\}$. Then h is a Lie subalgebra of g. For $X \in h$ implies $sX \in h$ for any $s \in R$. Next, for $X, Y \in h$, we have from the formulas (f) and (g) of remark (1) following Theorem 5.16 that for any $t \in R$, $\exp t(X + Y) \in H$ since H is closed. Thus $X + Y \in h$ and similarly $[XY] \in h$ so that h is a Lie subalgebra of g.

(b) Let m be a subspace of g such that $g = m + h$ (subspace direct sum). Then there is a neighborhood D of 0 in m such that $0 \neq X \in D$ implies $\exp X \notin H$. Suppose this is false. Then there is a sequence $\{Y_i\} \subset m$ such that $\lim Y_i = 0$ and $\exp Y_i \in H$. Now we can regard m as a Euclidean space with norm $\| \ \|$ and let $k = \{X \in m : 1 \leq \|X\| \leq 2\}$. Then we can choose integers n_i such that $X_i = n_i Y_i \in k$. Thus since k is compact, we can assume that $0 \neq X = \lim X_i$ exists in $k \subset m$ (passing to a subsequence if necessary). Furthermore, since $\lim Y_i = 0$, we see that $\lim 1/n_i = 0$ and $\exp(1/n_i X_i) = \exp Y_i \in H$.

We shall now show that the above X is in h and obtain the contradiction $X \in m \cap h = \{0\}$. Thus let $t_i = 1/n_i$ so that $\lim t_i = 0$. Then since $\exp(-t_i X_i) = (\exp t_i X_i)^{-1} \in H$, we can assume $t_i > 0$. Similar reasoning shows that $X \in h$ if $\exp tX \in H$ for all $t > 0$. Now for each $t > 0$, let

$$k_i(t) = \text{largest integer} \leq t/t_i.$$

Then since $(t/t_i) - 1 < k_i(t) \leq t/t_i$ and $\lim t_i = 0$, we see

$$\lim_i k_i(t)t_i = t.$$

Since $\exp t_i X_i \in H$, we see $\exp k_i(t)t_i X_i = [\exp t_i X_i]^{k_i(t)} \in H$. However, since exp is continuous and $\lim k_i(t)t_i X = tX$, we obtain

$$\exp tX = \lim \exp k_i(t)t_i X_i \in H.$$

Thus by definition of h in (a) we have the contradiction $X \in h$.

(c) For some neighborhood U in G with $e \in U$, we have

$$U \cap H = U \cap \exp h = \exp V$$

for some neighborhood V of 0 in h. For using the coordinates of the second kind relative to $g = h + m$, we can find neighborhoods V (respectively V') of 0 in h (respectively m) and U of e in G such that for $W = V + V'$ the map

$$\phi : W \to U : X + X' \to \exp X \exp X'$$

is a diffeomorphism. Now for $x \in U = \phi(W)$, we have $x = \exp X \exp X'$ with $X \in V \subset h$ and $X' \in V' \subset m$. Thus if $x \in U \cap H$, then $x = \exp X \exp X' \in H$ implies $\exp X' \in H$ since H is a group. Thus from (b) (shrinking V' to D if necessary) we see that $X' = 0$; that is, $U \cap H = U \cap \exp h = \exp V$.

(d) The set H is an analytic submanifold of G. To show that H is the underlying set of a submanifold, we use the converse of Proposition 2.11. Thus for $p \in H$, we must show that there is a neighborhood U in G with $p \in U$ such that $U \cap H$ is a submanifold of U; that is, there is a coordinate function z so that (U, z) is a coordinate system of G at p which satisfies (a) and (b) of Proposition 2.11. However, we can multiply by p^{-1} and translate the situation to $e \in G$. But these results follow from (c) above when we take $z_1, \ldots, z_m, \ldots, z_n$ to be canonical coordinates on U as given in (c), and we take the neighborhood W (in Proposition 2.11) to be $\exp V = U \cap H$. Also because U is open in G, this formula yields the topology, induced by the analytic structure, equals the topology induced on H by G. With this induced topology, we see that the map $R \to H : t \to \exp tX$ for $X \in h$ is continuous, and therefore from Proposition 6.4, h equals the Lie algebra of H.

Remark (1) Let H be a closed Lie subgroup with Lie algebra h as given above, and let $\varepsilon > 0$ be given. Then

$$h = \{X \in g : \exp sX \in H \text{ for all } s \in R \text{ with } |s| < \varepsilon\}.$$

For any given $t \in R$, there exists m so that $t = \sum_{i=1}^{m} s_i$, where $|s_i| < \varepsilon$ and we have $\exp tX = \exp s_1 X \cdots \exp s_m X \in H$.

Definition 6.10 Let A denote a finite-dimensional nonassociative algebra over R. Thus A is a finite-dimensional vector space with a bilinear map $\alpha : A \times A \to A$. Let B be another finite-dimensional (nonassociative) algebra with multiplication function β.

(a) A linear transformation $F : A \to B$ is an **algebra homomorphism** if $F(\alpha(X, Y)) = \beta(F(X), F(Y))$ for all $X, Y \in A$.

(b) The **kernel** of an algebra homomorphism $F : A \to B$ is the set $\ker F = \{X \in A : F(X) = 0\}$.

(c) A subspace D of an algebra A is an **ideal** if for all $X \in A$,

$$\alpha(X, D) = \{\alpha(X, Z) : Z \in D\} \subset D$$

and

$$\alpha(D, X) = \{\alpha(Z, X) : Z \in D\} \subset D.$$

Exercise (1) Analogous to the case for associative algebras show: D is an ideal of A if and only if D is the kernel of some algebra homomorphism $F : A \to B$.

Proposition 6.11 Let G and H be Lie groups with Lie algebras g and h and let $f : G \to H$ be a Lie group homomorphism.

(a) $\operatorname{Ker} f$ is a closed normal Lie subgroup of G and the Lie algebra of $\operatorname{Ker} f$ equals $\ker(Tf(e))$ which is an ideal in g where $Tf(e) : g \to h$ is the corresponding Lie algebra homomorphism.

(b) If G is connected, then $\operatorname{Im}(f)$, the image of f, is a Lie subgroup of H and the Lie algebra of $\operatorname{Im}(f)$ equals $\operatorname{Im}(Tf(e))$.

PROOF (a) Let $\bar{e}$ be the identity element in H. Then $\operatorname{Ker}(f) = \{x \in G : f(x) = \bar{e}\} = f^{-1}(\bar{e})$ is a normal subgroup which is closed in G. Thus by Theorem 6.9, $\operatorname{Ker}(f)$ is a Lie subgroup and its Lie algebra

$$\mathscr{L}(\operatorname{Ker} f) = \{X \in g : f(\exp tX) = \bar{e} \text{ for all } t \in R\}.$$

Thus $X \in \mathscr{L}(\operatorname{Ker} f)$ if and only if $\bar{e} = f(\exp_G tX) = \exp_H(tTf(e)X)$ for all $t \in R$. This implies $Tf(e)X = 0$ using exp as injective for t near enough $0 \in R$; that is, $\mathscr{L}(\operatorname{Ker} f) \subset \ker(Tf(e))$. From the same formulas, the other inclusion is clear. Also by the preceding exercise, $\ker(Tf(e))$ is an ideal.

For (b) note that since $Tf(e)$ is a Lie algebra homomorphism, $k = \operatorname{Im}(Tf(e))$ is a Lie subalgebra of h. Thus let K be the connected Lie subgroup of H with Lie algebra k. Then K is generated by elements $\exp(Tf(e)X)$ with $X \in g$. However, $\operatorname{Im}(f) = f(G)$ is generated by elements $f(\exp X)$ for $X \in g$, since G is connected. Therefore since $f(\exp X) = \exp(Tf(e)X)$, we have $K = f(G)$ since both of these subgroups are connected.

We shall consider the corresponding quotient groups in the next section on homogeneous spaces.

Exercises (2) Let G and H be connected Lie groups and let $f: G \to H$ be a Lie group homomorphism and let $Tf(e): g \to h$ be the corresponding Lie algebra homomorphism. Then show:

(i) f is surjective if and only if $Tf(e)$ is surjective;
(ii) f is injective implies $Tf(e)$ is injective. What about the converse?

(3) Let G be a Lie group and H an open subset which is a subgroup of G. Then show H is a Lie subgroup of G such that $G_0 = H_0$ (identity components).

REMARK (2) Using the results of this chapter, we briefly review the subgroups $SL(n)$, $SO(n)$, etc. of $GL(n, R)$ discussed in Section 2.3. Thus let R^* be the multiplicative (Lie) group of the nonzero real numbers. Then $\det: GL(n, R) \to R^*$ is an analytic homomorphism of Lie groups. From this $SL(n) = \operatorname{Ker} \det = \{A \in GL(n, R) : \det A = 1\}$ is a closed subgroup of $GL(n, R)$ and consequently a Lie subgroup (see Proposition 6.11). From $\det(\exp A) = e^{\operatorname{tr} A}$ we see $\det \circ \exp = \exp \circ \operatorname{tr}$ and using the chain rule we obtain $T(\det)(I) = \operatorname{tr}$. Thus from Proposition 6.11 the Lie algebra of Ker det equals $\ker T(\det)(I) = \ker \operatorname{tr} = \{A \in gl(n, R) : \operatorname{tr} A = 0\} = sl(n)$.

Let V be an n-dimensional vector space over R and let $B: V \times V \to R$ be a nondegenerate bilinear form on V. For $G = GL(V)$ let

$$H = \{A \in G : B(AX, AY) = B(X, Y) \quad \text{for all} \quad X, Y \in V\}.$$

Then H is a closed subgroup of G. Thus by Theorem 6.9, H is a Lie subgroup of G. Let h be the Lie algebra of H. Then, from Theorem 6.9, $A \in h$ if and only if $B((\exp tA)X, (\exp tA)Y) = B(X, Y)$ for all $t \in R$ and $X, Y \in V$. Thus as before we obtain

$$h = \{A \in gl(V) : B(AX, Y) + B(X, AY) = 0 \quad \text{for all} \quad X, Y \in V\}.$$

In particular, when B is positive definite and symmetric, we obtain the orthogonal group $H = O(n)$. Next since $f = \det: O(n) \to R^*$ is an analytic homomorphism we have the special orthogonal group $SO(n) = \operatorname{Ker} f$. Thus if superscript ^{t} denotes the transpose, the Lie algebra of $SO(n)$ equals $\ker Tf(I) = \{A \in gl(n) : A^t = -A\} = so(n)$. As before $\dim SO(n) = n(n-1)/2$ and since $SO(n)$ is connected we have Theorem 6.6 that $SO(n)$ is the unique connected Lie group generated by $\exp so(n)$. Recall from exercise (5). Section 3.2 that $O(n)$ is compact. Thus $SO(n)$ is compact since it is closed in $O(n)$.

Similarly the other groups $SO(p, q)$ and $Sp(p, R)$ are Lie groups with Lie algebras $so(p, q)$ and $sp(p, R)$ as discussed in Section 2.3. Are these groups compact?

4. Homogeneous Spaces

In this section we consider the coset space G/H and make it into an analytic manifold so that the projection $\pi : G \to G/H$ is analytic and the action of G on G/H given by $G \times G/H \to G/H : (a, xH) \to axH$ is also analytic. From a Lie algebra decomposition $g = m + h$ we use m to coordinatize G/H and give examples. Then we apply these results to quotient groups.

Thus let H be a closed subgroup of G so that H is a topological Lie subgroup according to Theorem 6.9. Consequently from Theorem 3.11, G/H is a Hausdorff space relative to the quotient topology and the natural projection $\pi : G \to G/H$ is open and continuous. Now let g and h be the Lie algebras of G and H, and let m be a subspace of g so that $g = m + h$ (direct sum). Then we use the canonical coordinates of the second kind to prove the following result [Helgason, 1962, p, 113].

Lemma 6.12 There exists a compact nucleus D of 0 in m and there exists a compact set N containing a neighborhood of $\bar{e} = eH$ in G/H such that $\exp|m : D \to \exp(D)$ is a homeomorphism, and $\pi : \exp(D) \to N$ is a homeomorphism. Thus $\pi \circ \exp|m : D \to N$ is a homeomorphism.

Proof As in the proof of Theorem 6.9 we use canonical coordinates of the second kind. Thus relative to $g = m + h$ (direct sum) we can find neighborhoods W (respectively W') of 0 in m (respectively h) and a neighborhood U of e in G such that

$$\phi(W + W') \cap H = U \cap H = \exp W',$$

where $U = \phi(W + W')$ is a coordinate neighborhood of e in G and ϕ is given by $\phi(X + X') = \exp X \exp X'$. Now let $\psi = \exp|m$ and let $D \subset W$ be a compact nucleus of 0 in m such that $\exp(-D)\exp D \subset U$. Then the restriction $\psi : D \to \psi(D)$ is a homeomorphism.

Set $N = \pi(\psi(D))$. We now show $\pi : \psi(D) \to N$ is injective. For let $X, Y \in D$ and assume $\pi(\exp X) = \pi(\exp Y)$. Then $(\exp X)H = (\exp Y)H$ so that

$$\exp(-X)\exp Y \in H \cap U = \exp W'.$$

Therefore there is $Z \in W'$ so that

$$\exp X \exp Z = \exp Y \exp 0.$$

However, since we are using coordinates of the second kind we have that $\phi | W + W'$ is injective. Thus $Z = 0$ which implies $\pi : \psi(D) \to N$ is injective. Thus $\pi|\psi(D)$ is a continuous injective map of the compact set $\psi(D)$ onto N and consequently a homeomorphism.

Note N is compact but it also contains an open neighborhood of $\bar{e}$ as follows. Since D contains an open neighborhood of 0 in m, $D + W'$ contains a neighborhood of $(0, 0)$ in $W + W'$. Therefore $\exp D \exp W'$ contains an open neighborhood of e in G and since $\pi : G \to G/H$ is an open map, $\pi(\psi(D)) = \pi(\exp D \exp W')$ contains an open neighborhood of $\bar{e}$ in G/H.

REMARKS (1) Let N^0 be the interior of the set $N = \pi(\psi(D))$, let D^0 be the interior of D, and let $k = \pi \mid \psi(D)$. Then for a fixed basis $X_1, \ldots, X_r$ of m the map

$$\psi^{-1} \circ k^{-1} : N^0 \to D^0 : \pi(\exp \textstyle\sum x_i X_i) \to \sum x_i X_i$$

is a homeomorphism which defines a chart (N^0, x) at $\bar{e} \in G/H$ where $x = \psi^{-1} \circ k^{-1}$. Consequently we can use this to define an analytic structure (on N^0).

(2) Using the coordinates of the second kind as above we see the map

$$g \to G : X + X' \to \exp X \exp X'$$

is an analytic diffeomorphism of a suitable neighborhood of 0 onto $E = \exp D^0 \exp W'$. Thus the map $\sigma : E \to m : \exp X \exp X' \to X$ is analytic. However, on E we see $\pi = x^{-1} \circ \sigma$ is analytic where x is the coordinate map above.

Next, using $N^0 = \pi(E)$ with E as above, we have for $p = \exp X \exp X' \in E$ that $x : N^0 \to m : pH \to X$. Thus the map $r = \exp \circ x : N^0 \to G$ is analytic. Furthermore $(\pi \circ r)(pH) = (\pi \circ \exp \circ x)(pH) = \pi(\exp X) = pH$; that is, $\pi \circ r = \text{identity} \mid N^0$. Combining these results we obtain that there exist neighborhoods E of e in G and N^0 of $\bar{e}$ in G/H so that:

(i) $\pi : E \to N^0$ is analytic;
(ii) there is an analytic map $r : N^0 \to G$ so that $\pi \circ r = \text{identity} \mid N^0$.

We apply these results to show the following.

Theorem 6.13 Let G be a Lie group, let H be a closed (Lie) subgroup, and let G/H have the quotient topology as before.

(a) Then G/H has a unique analytic structure such that:

(i) the projection map $\pi : G \to G/H$ is analytic;
(ii) every $\bar{p} \in G/H$ has a neighborhood P and an analytic map $\bar{r} : P \to G$ so that $\pi \circ \bar{r} = \text{identity} \mid P$.

(b) With the analytic structure in (a) on G/H we have for every $a \in G$ that the map $\tau(a) : G/H \to G/H : xH \to axH$ is an analytic diffeomorphism.

(c) With the analytic structure in (a) on G/H the map $\phi : G \times G/H \to G/H : (a, xH) \to axH$ is analytic.

PROOF First we assume (a) holds and prove (b). Since $\tau(a)$ has inverse $\tau(a^{-1})$ it suffices to show $\tau(a)$ is analytic. Then $\tau(a^{-1})$ is also analytic. We now note for $x \in G$ that

$$\tau(a) \circ \pi(x) = axH = \pi \circ L(a)(x)$$

so that on G we have $\tau(a) \circ \pi = \pi \circ L(a)$. Therefore for $\bar{x} = xH \in P$ as above, we see $\tau(a)(\bar{x}) = \tau(a) \circ (\text{identity})(\bar{x}) = \tau(a) \circ \pi \circ r(\bar{x}) = \pi \circ L(a) \circ r(\bar{x})$. Thus on P we see $\tau(a) = \pi \circ L(a) \circ r$ is a composition of analytic maps and therefore is analytic.

This implies that it is sufficient to prove the existence and uniqueness of the analytic structure at the point $\bar{e} = eH$ in G/H, for if a coordinate system exists on a neighborhood N^0 of $\bar{e}$, then we obtain a coordinate system at any other point $\bar{p} = pH$ by translation via the analytic map $\tau(p)$. Futhermore if (i) and (ii) hold for N^0, then they hold for $\tau(p)N^0 = P$. Briefly, if r satisfies (ii) for N^0, then $\bar{r} = L(p) \circ r \circ \tau(p^{-1})$ satisfies (ii) for P using $\tau(p) \circ \pi = \pi \circ L(p)$. The proof is similar for (i).

Now from the preceding remarks we have shown the existence of an analytic structure at $\bar{e}$ satisfying (i) and (ii) of (a) and we prove uniqueness. Suppose N^0 has two analytic structures with two coordinate maps x_1 and x_2 and maps r_1 and r_2 from E into G satisfying (a) where E is given in remark (2). We shall show $x_1 \circ x_2^{-1}$ is analytic. Thus $x_1 \circ x_2^{-1} = x_1 \circ \pi \circ r_2 \circ x_2^{-1}$. However from (i), $x_1 \circ \pi$ is an analytic map of G into some Euclidean space and from (ii), $r_2 \circ x_2{}^{-1}$ is analytic. Similarly $x_2 \circ x_1{}^{-1}$ is analytic so that from Section 2.2 we see that the analytic structures are equivalent.

To show $\phi : G \times G/H \to G/H : (a, xH) \to axH$ is analytic, it suffices to show that its restriction to $G \times N^0$ is analytic. Thus for r as defined in remark (2) above and for $\mu : G \times G \to G$ the multiplication function on G, we see $\phi | G \times N^0$ can be factored as follows.

$$G \times N^0 \xrightarrow{i_G \times r} G \times G \xrightarrow{\mu} G \xrightarrow{\pi} G/H,$$

and since these are analytic, so is $\phi = \pi \circ \mu \circ (i_G \times r)$.

REMARKS (3) Let G be a Lie group and let M be an analytic manifold such that G is a topological transformation group on M (Section 3.2). Then G is a **Lie transformation group** of M if the map $G \times M \to M : (a, x) \to ax$ is analytic. Thus in the above theorem G is a Lie transformation group of G/H. Furthermore, it is proved by Helgason [1962, p. 113] and Tondeur [1965, p. 155] that the above analytic structure on G/H is the unique analytic structure such that G (with the above action) is a Lie transformation group on G/H.

(4) For $p \in G$, the coordinates at $\bar{p} = pH$ as discussed in the above proof are explicitly given as follows. Let (N^0, x) be the coordinates at e given in remark (1) where

$$x : N^0 \to D^0 : \pi(\exp \sum x_i X_i) \to \sum x_i X_i .$$

Then for the neighborhood $P = \tau(p)N^0$ we have that

$$y : P \to D^0 : \pi(p \exp \sum x_i X_i) \to \sum x_i X_i$$

is a coordinate map where we regard $m = R^n$.

Exercise (1) Let H be a closed subgroup of the Lie group G and let $g = m + h$ be a fixed decomposition of the Lie algebras as above. Show that the map $\pi : G \to G/H$ yields the epimorphism $T\pi(e) : T(G, e) \to T(G/H, \bar{e})$. Thus identifying $g = T(G, e)$ show that $\ker T\pi(e) = h$ and consequently $T(G/H, \bar{e}) \cong g/h \cong m$.

Example (1) Let $G = SO(n)$ and for $p + q = n$ let $H = SO(p)$ which can be considered as a subgroup G by regarding H as matrices of the form

$$\begin{bmatrix} I & 0 \\ 0 & A \end{bmatrix}.$$

where A is a $p \times p$ orthogonal matrix with $\det A = 1$ and I the $q \times q$ identity matrix. The Lie algebra of G is $g = so(n)$ which is the set $n \times n$ skew-symmetric matrices. Thus the Lie algebra $h = so(p)$ is given by the set of matrices

$$\begin{bmatrix} 0 & 0 \\ 0 & B_{22} \end{bmatrix}.$$

where B_{22} is a $p \times p$ skew-symmetric matrix and m is given by the obvious complementary set of matrices in g, namely

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & 0 \end{bmatrix}. \tag{$*$}$$

Thus for X given by $(*)$ in a suitable neighborhood of 0 in m, the coordinates at $\bar{e} \in G/H$ as discussed above are given by

$$(\exp X)H \to X.$$

Exercise (2) Let S^{n-1} be represented by $SO(n)/SO(n-1)$ as in example (4), Section 3.2. Compare the coordinates near the point $\bar{I} \in SO(n)/SO(n-1)$ as given above with the stereographic projection coordinates given in example (5), Section 2.1.

Definition 6.14 Let A be a nonassociative algebra with multiplication function α and let D be an ideal of A. Then the coset vector space $\bar{A} = A/D = \{X + D : x \in A\}$ is made into an algebra as follows. For $\bar{X} = X + D$ and $\bar{Y} = Y + D$ in $\bar{A}$, define $\bar{\alpha} : \bar{A} \times \bar{A} \to \bar{A}$ by

$$\bar{\alpha}(\bar{X}, \bar{Y}) = \alpha(X, Y) + D = \overline{\alpha(X, Y)}.$$

The algebra $\bar{A}$ is called a **quotient algebra**.

As in the associative case, it is easy to verify that $\bar{\alpha}$ is a function which is bilinear so that $\bar{A}$ with multiplication $\bar{\alpha}$ becomes an algebra. Furthermore the map $A \to \bar{A} : X \to \bar{X}$ is an algebra homomorphism.

Proposition 6.15 Let G be a Lie group with Lie algebra g and let H be a closed normal (Lie) subgroup with Lie algebra h. Then the factor group G/H with the analytic structure given in Theorem 6.13 is a Lie group and its Lie algebra is isomorphic to the quotient algebra g/h where h is an ideal of g.

PROOF The factor group G/H is a topological group relative to the quotient topology and is an analytic manifold relative to the analytic structure of Theorem 6.13. Thus we must show that the operation

$$G/H \times G/H \to G/H : (xH, yH) \to xy^{-1}H$$

is analytic. However, by remark (2), Section 4.1 and the uniqueness of the analytic structure of a Lie group, it suffices to show the multiplication

$$\bar{\mu} : G/H \times G/H \to G/H : (xH, yH) \to xyH$$

is analytic at $(\bar{e}, \bar{e})$. Thus let N^0 be the neighborhood of $\bar{e}$ in remark (2) with $\pi \circ r = \text{identity} | N^0$ where $r : N^0 \to G$ is analytic. Then for $\bar{x}, \bar{y} \in N^0$ we use $\pi : G \to G/H$ as a group homomorphism to obtain

$$\bar{\mu}(\bar{x}, \bar{y}) = \bar{\mu}(\pi \circ r(\bar{x}), \pi \circ r(\bar{y})) = \pi\mu(r(\bar{x}), r(\bar{y})),$$

where μ is the multiplication in G. Thus $\bar{\mu} = \pi \circ \mu \circ (r \times r)$ which is analytic at $(\bar{e}, \bar{e})$.

Now $\pi : G \to G/H$ is an analytic homomorphism of Lie groups and if $\mathscr{L}(G/H)$ is the Lie algebra of G/H, then $T\pi(e) : g \to \mathscr{L}(G/H)$ is a Lie algebra homomorphism. Thus since Ker $\pi = H$, we have by Proposition 6.11 and the characterization of h in Theorem 6.9 that $\ker(T\pi(e)) = h$ is an ideal. Thus $\mathscr{L}(G/H)$ is isomorphic to g/h.

Corollary 6.16 Let $f : G \to H$ be a Lie group homomorphism and assume that G is connected. Let $\bar{f} : G/\text{Ker}\, f \to f(G)$ be the induced isomorphism as abstract groups. Then $\bar{f}$ is an isomorphism of Lie groups where the Lie group structure of $f(G)$ (respectively $G/\mathrm{Ker} f$) is given by Proposition 6.11 (respectively Theorem 6.13).

Exercise (3) Let $f: G \to H$ be a surjective analytic homomorphism of the Lie groups G and H. Let H' be a closed normal subgroup of H and let $G' = f^{-1}(H')$. Then show that the isomorphism $G/G' \cong H/H'$ of abstract groups is actually a Lie group isomorphism.

5. Commutative Lie Groups

We shall apply some of the preceding results on homomorphisms to show that a connected commutative (i.e., Abelian) Lie group G is isomorphic to $R^q \times T^p$ for suitable integers p and q. This also shows that a connected Lie group G is commutative if and only if its Lie algebra g is commutative.

Recall (Section 3.1) that a topological subgroup H of a topological group G is a discrete subgroup if H is a discrete subspace of G; that is, every subset of H is open in H. If G is a Lie group, then a discrete subgroup H can be regarded as a zero-dimensional Lie group and in this case H is a closed subgroup of G.

Proposition 6.17 The set H is a discrete subgroup of the additive Lie group R^n if and only if H is isomorphic to $Z^p = Z \times \cdots \times Z$ (p-times) for some p with $0 \leq p \leq n$.

PROOF It is clear that Z^p is a discrete subgroup of R^n. Conversely, we shall now show there are linearly independent elements $u_1, \ldots, u_p$ of R^n which generate the discrete group H as a Z-submodule of R^n. Thus $H = Zu_1 + \cdots + Zu_p$ which gives the result.

Lemma 6.18 Let H be a nonzero discrete subgroup of R^n. Then H is generated as a group by elements which are linearly independent in R^n.

PROOF We show this by induction on the dimension n. Thus for $n = 0$, the result follows. Now for $n \geq 1$, assume the lemma is true for $n - 1$ and let H be a nonzero discrete subgroup of R^n. Let $\{u_1, \ldots, u_m\}$ be a maximal set of linearly independent elements (over R) of H. First assume $m < n$. Then for $n = 1$ this set is empty. Thus for $n > 1$, we see that H is contained in a proper subspace; that is, $H \subset Ru_1 + \cdots + Ru_m = R^m$ for $m < n$. We can now apply the induction hypothesis to obtain the result.

Next assume $m = n$ and let V be the subspace of R^n generated by $\{u_1, \ldots, u_{n-1}\}$. Then $H \cap V$ is a discrete subgroup of V and so by the induction hypothesis we can assume $H \cap V$ is generated by $\{u_1, \ldots, u_{n-1}\}$ as a Z-module.

Let $Q = \{\sum_{i=1}^n b_i u_i : 0 \le b_i \le 1 \text{ for } i = 1, \ldots, n\}$ be the compact cube in R^n. Since H is discrete, Q contains only finitely many elements of H. Let

$$P = \left\{\sum_{i=1}^n b_i u_i \in Q \cap H : b_n > 0\right\},$$

Then $u_n \in P$ and let u be an element in P with b_n minimal (since there are only finitely many elements in P). We shall now show that the basis $\{u_1, \ldots, u_{n-1}, u\}$ of R^n generates H as a group; that is, $H = Zu_1 + \cdots + Zu_{n-1} + Zu$. Thus let $h = \sum_{i=1}^{n-1} h_i u_i + h_n u \in H$ for $h_i \in R$ with $i = 1, \ldots, n$. Now we shall show $h_n \in Z$. Then since $h - h_n u \in H \cap V$, we have by the results of the preceding paragraph that $h_1, \ldots, h_{n-1} \in Z$. Let $u = \sum_{i=1}^n b_i u_i \in P$ (with b_n minimal) and let k_n be the largest integer smaller than h_n and let k_i be the largest integer smaller than $h_i + b_i(h_n - k_n) \equiv c_i$ for $i = 1, \ldots, n-1$. The following element in H

$$\begin{aligned}
h - \sum_{i=1}^{n-1} k_i u_i - k_n u &= \sum_{i=1}^{n-1} (h_i - k_i)u_i + (h_n - k_n)u \\
&= \sum_{i=1}^{n-1} [c_i - b_i(h_n - k_n) - k_i]u_i + (h_n - k_n) \sum_{i=1}^{n} b_i u_i \\
&= \sum_{i=1}^{n-1} (c_i - k_i)u_i + (h_n - k_n)b_n u_n
\end{aligned}$$

is in Q because by definition

$$0 < c_i - k_i < 1$$

and by definition $0 < h_n - k_n < 1$ yields

$$0 < b_n(h_n - k_n) < b_n \le 1.$$

Therefore this element is in P. However, this last equation also contradicts the definition of b_n as minimal in the choice of $u \in P$ if we assume h_n is not an integer. Thus we must have $h_n \in Z$.

Proposition 6.19 Let G be a commutative Lie group with Lie algebra g. Then, regarding g as an additive Lie group R^m, the map $\exp : g \to G$ is a homomorphism of Lie groups and Ker exp is a discrete subgroup of g.

PROOF From Section 5.3 we have seen $\exp : g \to G$ is a homomorphism. Now there exist neighborhoods U_0 of 0 in g and U_e of e in G such that $\exp | U_0 : U_0 \to U_e$ is a diffeomorphism. Therefore, since exp is injective on U_0,

$$(\text{Ker exp}) \cap U_0 = \{0\}.$$

Thus since U_0 is open in g, the set $\{0\}$ is open in Ker exp. However, translations are homeomorphisms so that $\{a\}$ is open for every $a \in$ Ker exp. Thus since any set W of Ker exp is of the form $W = \bigcup_{a \in W} \{a\}$, W is open; that is, Ker exp is discrete.

Theorem 6.20 Let G be a connected commutative Lie group of dimension n. Then there exists an integer p, $0 \leq p \leq n$, so that G is isomorphic to $R^{n-p} \times T^p$.

PROOF Since $\exp X \exp Y = \exp(X + Y)$ and the connected group G is generated by a neighborhood $U = \exp U_0$, we see that the homomorphism $\exp : g \to G$ is surjective (also Section 5.3). From Corollary 6.16 (or directly), we have G and $g/\text{Ker exp}$ are isomorphic as Lie groups. However, regarding g as the additive group R^n, we have from the preceding results Ker exp is isomorphic to some Z^p. Thus G is isomorphic to $R^n/Z^p = (R^{n-p} \times R^p)/Z^p$ which gives the results since T^p is isomorphic to R^p/Z^p extending example (1), Section 3.2.

Exercise (1) The results outlined below will be used in the chapter on solvable Lie groups.

(i) Let G be a commutative group with $+$ as its operation. Then G is called **divisible** if for any integer $n \in Z$ and any $x \in G$ we have $x \in nG = \{ny : y \in G\}$. Show if G and H are commutative divisible groups, then the direct sum $G \oplus H \cong G \times H$ is also divisible. Show that the additive groups R and $T^1 = R/Z$ are divisible. Thus $R^q \times T^p$ is divisible.

(ii) Let H be a subgroup of the commutative group G. Then H is a **divisible subgroup** if H is divisible as a group. Show that if H is a divisible subgroup of a commutative group G, then there exists a subgroup K of G such that $G = H \oplus K$. (This is not too easy [MacLane, 1963; Barns, 1965].)

(iii) Now let G be a commutative Lie group (written additively). Then we know that the identity component $G_0 \cong R^q \times T^p$ is divisible. Therefore we have an exact sequence

$$0 \longrightarrow G_0 \xrightarrow{\ i\ } G \xrightarrow{\ \pi\ } G/G_0 \longrightarrow 0$$

which splits since $G = G_0 \oplus K$. Because of this $G \xrightarrow{\pi} G/G_0 \to 0$ splits and therefore there exists a Lie group homomorphism $f : G/G_0 \to G$ with $\pi \circ f$ the identity on G/G_0. Using this and the fact that G/G_0 is discrete (Theorem 3.22), show that K is a discrete Lie subgroup of G. Thus we have shown that if G is a commutative Lie group (written additively), then $G = G_0 \oplus K$ where the connected component $G_0 \cong R^q \times T^p$ is a divisible subgroup and K is a discrete subgroup.

CHAPTER 7

AUTOMORPHISMS AND ADJOINTS

We have already considered some results on automorphisms of Lie groups; for example, an automorphism $\phi : G \to G$ induces an automorphism $T\phi(e) : g \to g$ of the corresponding Lie algebra. We now develop some of the "structural" results for groups of automorphisms. Thus we first show that the automorphism group of a nonassociative algebra A is a Lie group whose Lie algebra is the derivation algebra of A. Next we develop the concept of inner derivations of a nonassociative algebra and the corresponding inner automorphisms. When the algebra is associative these reduce to the usual concepts of inner derivations and automorphisms. Using these results we use the differential of an inner automorphism $\phi(a) : G \to G : x \to axa^{-1}$ of a Lie group G to obtain an automorphism $(T\phi(a))(e) : g \to g$ of the corresponding Lie algebra g. Then we obtain a mapping $\mathrm{Ad} : G \to GL(g) : a \to (T\phi(a))(e)$ called the "adjoint representation of G." We develop formulas for the adjoint representation which lead to the result that the inner automorphism group of a connected Lie group G equals $\mathrm{Ad}(G)$. The fact that for a connected Lie group G, $\mathrm{Aut}(G)$ is a Lie group is proved in Chapter 8.

1. Automorphisms of Algebras

Let A be a finite-dimensional nonassociative algebra over a field K with bilinear multiplication function α and define $\alpha(X, Y) = X \cdot Y = XY$. Let $\mathrm{Aut}(A)$ denote the automorphism group of A; that is, the set of maps $\phi \in GL(A)$ with $\phi(XY) = \phi(X)\phi(Y)$. Then $\mathrm{Aut}(A)$ is a subgroup of $GL(A)$.

Proposition 7.1 Let A be a nonassociative algebra over R. Then $\mathrm{Aut}(A)$ is a closed subgroup of $GL(A)$. Thus $\mathrm{Aut}(A)$ is a closed Lie subgroup of $GL(A)$.

PROOF For X, Y fixed in A we note that since the multiplication in A is continuous (it is a bilinear map of the finite-dimensional space A), the set $S(X, Y) = \{\phi \in GL(A) : \phi(XY) = \phi(X)\phi(Y)\}$ is closed. Thus $\mathrm{Aut}(A) = \bigcap \{S(X, Y) : X, Y \in A\}$ is closed in $GL(A)$.

Definition 7.2 A **derivation** D of a nonassociative algebra A is a linear transformation of A satisfying $D(X \cdot Y) = DX \cdot Y + X \cdot DY$ for all $X, Y \in A$.

Let $\mathrm{Der}(A)$ denote the set of derivations of A. Then it is easy to see that $\mathrm{Der}(A)$ is a Lie subalgebra of $gl(A)$.

Proposition 7.3 Let D be a derivation of a nonassociative algebra A over R. Then:

(a) $\exp tD \in \mathrm{Aut}(A)$ for all $t \in R$;
(b) the Lie algebra of $\mathrm{Aut}(A)$ equals $\mathrm{Der}\,(A)$.

PROOF Since $\mathrm{Aut}(A)$ is a closed subgroup of $GL(A)$, (a) follows from (b) and Theorem 6.9. So to show (b), let D be in the Lie algebra of $\mathrm{Aut}(A)$. Then from Theorem 6.9, we have $\exp tD \in \mathrm{Aut}(A)$ for all $t \in R$. Therefore for $X, Y \in A$

$$(\exp tD)(X \cdot Y) = (\exp tD)X \cdot (\exp tD)Y$$

and using the product rule, we differentiate with respect to t at $t = 0$ to obtain $D(X \cdot Y) = DX \cdot Y + X \cdot DY$; that is, $D \in \mathrm{Der}(A)$. To show the converse, we let D be a derivation and let

$$\alpha(t) = (\exp tD)(XY) \qquad \text{and} \qquad \beta(t) = (\exp tD)X \cdot (\exp tD)Y.$$

Then we see $\alpha(0) = \beta(0) = XY$ and $\dot{\alpha}(t) = D\alpha(t)$ in A. Also using D as a derivation we have

$$\begin{aligned} \dot{\beta}(t) &= D(\exp tD)X \cdot (\exp tD)Y + (\exp tD)X \cdot D(\exp tD)Y \\ &= D[(\exp tD)X \cdot (\exp tD)Y] = D\beta(t). \end{aligned}$$

Thus α and β are solutions to $\dot{z} = Dz$ satisfying the same initial conditions. This implies the desired result $\alpha(t) = \beta(t)$ using Proposition 2.36.

2. Inner Derivations and Automorphisms

Recall that for a nonassociative algebra A over R we have defined the mappings $L(X) : A \to A : Y \to XY$ and $R(X) : A \to A : Y \to YX$ for all $X \in A$. In particular, if $A = g$ a Lie algebra, we have set

$$L(X) = \mathrm{ad}(X)$$

called the **adjoint** mapping in g.

Definition 7.4 Let A be a nonassociative algebra over R and let P be the subspace of $\mathrm{End}(A)$ spanned by $L(X)$ and $R(Y)$ for all $X, Y \in A$. The **Lie transformation algebra of** A, denoted by $L(A)$, is the Lie subalgebra of $gl(A)$ generated by P [Schafer, 1966].

Examples (1) If A is a Lie algebra with multiplication $[XY]$, then $L(X) = \mathrm{ad}(X) = -R(X)$ so that $P = \{\mathrm{ad}(X) : X \in A\}$. Using the Jacobi identity, we see that

$$[\mathrm{ad}\, X, \mathrm{ad}\, Y] = \mathrm{ad}([XY])$$

so that P is a Lie algebra of linear transformations; that is, $L(A) = P$

(2) If A is associative, then we have

$$\begin{aligned} &L(X)L(Y)Z = X(YZ) = (XY)Z = L(XY)Z &&\text{and} && L(X)L(Y) = L(XY);\\ &R(X)R(Y)Z = (ZY)X = R(YX)Z &&\text{and} && R(X)R(Y) = R(YX);\\ &L(X)R(Y)Z = X(ZY) = R(Y)L(X)Z &&\text{and} && L(X)R(Y) = R(Y)L(X). \end{aligned}$$

Thus in this case P is also closed under commutation; that is, $L(A) = P$. We shall give examples later when $L(A)$ is more complicated.

Definition 7.5 A derivation D of a nonassociative algebra A is **inner** if $D \in L(A)$. Let $\mathrm{Inn}(A)$ denote the set of inner derivations of A.

Examples (3) If A is a Lie algebra, then for any $X \in A$ we have $D = \mathrm{ad}(X)$ is an inner derivation.

(4) Let A be associative, and let $D = L(X) + R(Y)$ be an inner derivation. If A has an identity element 1, then

$$D(1) = D(1 \cdot 1) = D(1) \cdot 1 + 1 \cdot D(1)$$

so that $0 = D(1) = (L(X) + R(Y))(1) = X + Y$. Thus for an associative algebra with identity, inner derivations are of the form $D = L(X) - R(X)$.

Proposition 7.6 The set $\mathrm{Inn}(A)$ of inner derivations of a nonassociative algebra A is an ideal in the Lie algebra $\mathrm{Der}(A)$.

PROOF We first find a formula for $L(A)$ as follows. Let $P_1 = P$ be as in the definition of $L(A)$ and let $P_{i+1} = [P_1, P_i]$. Then we have

$$[P_i, P_m] \subset P_{i+m} \qquad \text{for} \quad i, m = 1, 2, \ldots.$$

The case $i = 1$ follows from the definition. Now assume the results for $i = k$ and all m. Then we use the Jacobi identity in $gl(A)$ as follows for $i = k + 1$.

$$\begin{aligned}[P_{k+1}, P_m] &\subset [[P_1, P_k], P_m] \\ &\subset [[P_1, P_m], P_k] + [P_1, [P_k, P_m]] \\ &\subset [P_{m+1}, P_k] + [P_1, P_{k+m}] \\ &\subset P_{k+m+1},\end{aligned}$$

where we use the induction hypothesis on the third containment. Thus $\sum_{i=1}^{\infty} P_i$ is a Lie subalgebra of $L(A)$ which contains P so that by definition we have

$$L(A) = \sum_{i=1}^{\infty} P_i.$$

Next we note by induction that if $D \in \mathrm{Der}(A)$, then $[D, P_i] \subset P_i$. For $i = 1$, this uses the formulas

$$[D, L(X)] = L(DX) \qquad \text{and} \qquad [D, R(X)] = R(DX)$$

which follow from the definition of a derivation. For $i > 1$, we use the Jacobi identity again. Now since $L(A) = \sum P_i$, this yields

$$[\mathrm{Der}(A), L(A)] \subset L(A).$$

Finally, since $\mathrm{Inn}(A) = \mathrm{Der}(A) \cap L(A)$ is a subspace of $\mathrm{Der}(A)$, we see

$$[\mathrm{Inn}(A), \mathrm{Der}(A)] \subset L(A) \cap \mathrm{Der}(A) = \mathrm{Inn}(A);$$

that is, $\mathrm{Inn}(A)$ is an ideal in $\mathrm{Der}(A)$.

Definition 7.7 An automorphism ϕ of A is **inner** if ϕ is contained in the subgroup of $\mathrm{Aut}(A)$ generated by $\exp(\mathrm{Inn}(A))$. Thus $\phi = \exp D_1 \cdots \exp D_n$ where $D_i \in \mathrm{Inn}(A)$. Let $\mathrm{Int}(A)$ denote the subgroup of inner automorphisms of A.

Example (5) Let A be a finite-dimensional associative algebra of endomorphisms containing the identity I. Let $D = L(X) - R(X) \in \mathrm{Inn}(A)$. Then since $L(X)R(X) = R(X)L(X)$ we have

$$\begin{aligned}\exp D &= \exp(L(X) - R(X)) \\ &= \exp L(X) \exp(-R(X)) = \exp L(X) \exp R(-X)\end{aligned}$$

using $\exp(S+T) = \exp S \exp T$ if $ST = TS$. However, $[\exp L(X)](Z) = (\sum L(X)^n/n!)Z = (\sum X^n/n!)Z = (\exp X)Z$ and similarly, $[\exp R(X)]Z = Z(\exp X)$. Thus we obtain

$$\begin{aligned}(\exp D)Z &= [\exp L(X) \exp R(-X)]Z \\ &= (\exp X)Z(\exp -X) = (\exp X)Z(\exp X)^{-1}\end{aligned}$$

which conforms to the usual definition of an inner automorphism as an endomorphism $\phi \in GL(A)$ of the form $\phi(Z) = UZU^{-1}$ for some $U \in A$.

Later we shall give a criterion which shows that for many simple algebras, all derivations are inner.

3. Adjoint Representations

We now consider the differential of inner automorphisms to obtain a representation of a Lie group G in its Lie algebra g.

Definition 7.8 Let V be a finite-dimensional vector space over a field K, let G be a group, let g be a Lie algebra over K, and let A be an associative algebra over K. Then a **group representation of** G **in** V is a group homomorphism $G \to GL(V)$. A **Lie algebra representation of** g **in** V is a Lie algebra homomorphism $g \to gl(V)$. An **associative algebra representation of** A **in** V is an associative algebra homomorphism $A \to \mathrm{End}\ (V)$. An injective representation is called **faithful**.

Examples (1) If g is a Lie algebra over R and if $\mathrm{ad}(g) = \{\mathrm{ad}\ X : X \in g\}$, then $\mathrm{ad} : g \to \mathrm{ad}(g) : X \to \mathrm{ad}\ X$ is a Lie algebra representation called the **adjoint representation**. For in this case we use the Jacobi identity to obtain $\mathrm{ad}[XY] = [\mathrm{ad}\ X, \mathrm{ad}\ Y]$ so that the map ad is a homomorphism. Next note that $\ker(\mathrm{ad}) = \{Z \in g : [XZ] = 0 \text{ for all } X \in g\}$ is the **center** $Z(g)$ of g. Consequently the Lie algebras $\mathrm{ad}(g)$ and $g/Z(g)$ are isomorphic. Thus if $Z(g) = \{0\}$ then g is isomorphic to a Lie algebra of linear transformations.

(2) If A is an associative algebra, then the map $A \to \mathrm{End}(A) : X \to L(X)$ is a representation which is faithful if A has an identity element.

(3) Let G be a connected Lie group, and let $\phi \in \mathrm{Aut}(G)$ be an analytic automorphism. Then $T\phi(e) \in GL(g)$, and the map

$$\mathrm{Aut}(G) \to GL(g) : \phi \to T\phi(e)$$

is a faithful representation of the subgroup of analytic automorphisms of Aut(G) in g. [Note Theorem 8.14 for the Lie structure of Aut(G).] To see this is a representation, we note for analytic automorphisms ϕ, $\psi \in$ Aut(G),

$$T(\phi \circ \psi)(e) = T\phi(\psi(e)) \circ T\psi(e) = T\phi(e) \circ T\psi(e).$$

This representation is faithful; for if $(T\phi(e))(X) = I(X) = X$ for all $X \in g$, then by Proposition 5.19,

$$\phi(\exp X) = \exp(T\phi(e)(X)) = \exp X.$$

Thus $\phi = idy$ on a neighborhood U of e in G and since G is connected, G is generated by U. Thus since ϕ is an automorphism, it is the identity on G (or we can use Proposition 5.20). Since continuous automorphisms of G are actually analytic, this example extends to a representation of the group of continuous automorphisms of G.

Definition 7.9 Let G be a Lie group with Lie algebra g and for $a \in G$ let $\phi(a) : G \to G : x \to axa^{-1}$ be the corresponding (analytic) inner automorphism. Then the mapping

$$\mathrm{Ad} : G \to GL(g) : a \to (T\phi(a))(e)$$

is called the **adjoint representation of** G.

Proposition 7.10 The adjoint representation is an analytic homomorphism of G into the Lie group Aut(g).

PROOF Since $\phi(a)$ is an analytic automorphism of G, $(T\phi(a))(e)$ is an automorphism of g and

$$T\phi(ab)(e) = (T(\phi(a) \circ \phi(b)))(e) = T\phi(a)(e) \circ T\phi(b)(e).$$

Thus $\mathrm{Ad}(ab) = \mathrm{Ad}(a) \circ \mathrm{Ad}(b)$ so that Ad is a representation of G. To show that it is analytic, let $X_1, \ldots, X_m$ be a basis of g and let $\{x_1, \ldots, x_m\}$ be the corresponding canonical coordinate system. Now let

$$\mathrm{Ad}(a)(X_i) = [T\phi(a)(e)](X_i) = \sum_j a_{ji}(a)X_j$$

be the matrix representation of Ad(a). If $x = \exp tX_i$, then from Proposition 5.19

$$\begin{aligned}\phi(a)x &= \phi(a)(\exp tX_i) \\ &= \exp[tT\phi(a)(e)(X_i)] = \exp[t \sum_j a_{ji}(a)X_j].\end{aligned}$$

Thus for t near enough to $0 \in R$, $ta_{ji}(a)$ are the canonical coordinates of $\phi(a)x$. However, the canonical coordinates for $\phi(a)x$ are given by commutator formula in Theorem 5.16(b) and are analytic for a in a suitable neighborhood of e in G. Thus Ad is analytic at $e \in G$ and since it is a homomorphism of Lie groups, Ad is analytic on all of G.

We shall now derive some formulas for Ad.

Theorem 7.11 Let $\mathrm{Ad} : G \to \mathrm{Aut}(g)$ be the adjoint representation of a Lie group G. Then

$$(T\,\mathrm{Ad})(e) = \mathrm{ad};$$

that is, the differential at e of the adjoint representation of the Lie group is the adjoint representation of the Lie algebra.

Proof Since $\mathrm{Ad} : G \to \mathrm{Aut}(g)$ we have $T\,\mathrm{Ad}(e) : T(G, e) \to T(\mathrm{Aut}(g), I)$. Thus for $X \in g = T(G, e)$ we have the value $X^* = (T\,\mathrm{Ad})(e)X \in T(\mathrm{Aut}(g), I) = \mathrm{Der}(g)$; that is, $X^* \in \mathrm{End}(g)$. Since Ad is a homomorphism, we have for t near 0 in R

$$(\mathrm{Ad})(\exp tX) = \exp[t(T\,\mathrm{Ad})(e)X] = \exp tX^*. \tag{1}$$

Thus since $\mathrm{Ad}(\exp tX) \in \mathrm{End}(g)$, we have for $Y \in g$ and from the series $\exp tX^*$ for the linear transformation X^*,

$$\begin{aligned} X^*(Y) &= \lim_{t\to 0} \frac{1}{t} (\exp tX^* - I)(Y) \\ &= \lim_{t\to 0} \frac{1}{t} (\mathrm{Ad}(\exp tX) - I)(Y). \end{aligned} \tag{2}$$

Now for s and t sufficiently near 0 in R we have

$$\begin{aligned} \exp[s\,\mathrm{Ad}(\exp tX)Y] &= \exp[s(T\phi(\exp tX)(e))(Y)] \\ &= \phi(\exp tX)[\exp(sY)] \\ &= (\exp tX)(\exp sY)(\exp -tX) \\ &= \exp(sY + st[XY] + so(t^2)) \end{aligned}$$

using the definition of Ad, the fact $\sigma(\exp X) = \exp(T\sigma(e)X)$ for an analytic automorphism σ of G, and Theorem 5.16(b). Therefore we can conclude for t near 0 in R

$$\mathrm{Ad}(\exp tX)Y = Y + t[XY] + o(t^2). \tag{3}$$

Thus substituting (3) into (2) we obtain

$$X^*(Y) = \lim_{t\to 0} \frac{1}{t} (Y + t[XY] - Y) = [XY];$$

that is, $[T\,\mathrm{Ad}(e)](X) = X^* = \mathrm{ad}\,X$.

Corollary 7.12 For $X \in g$ and $a \in G$ we have

(a) $\mathrm{Ad}(\exp X) = \exp(\mathrm{ad}\, X)$ $(= e^{\mathrm{ad} X})$
(b) $a(\exp X)a^{-1} = \exp[\mathrm{Ad}(a)(X)]$.

PROOF Part (a) is just a restatement of formula (1) in the above proof. For (b) we have as in the above proof that for the inner automorphism $\phi(a) : G \to G : x \to axa^{-1}$,

$$\begin{aligned} a(\exp X)a^{-1} &= \phi(a)(\exp X) \\ &= \exp[(T\phi(a)(e))(X)] = \exp[\mathrm{Ad}(a)(X)] \end{aligned}$$

using the definition of Ad.

Corollary 7.13 If G is a connected Lie group with Lie algebra g, then $\mathrm{Ad}(G) = \{\mathrm{Ad}\, a : a \in G\}$ is a Lie group with Lie algebra $\mathrm{ad}(g)$. Thus $\mathrm{Int}(g)$ the inner automorphism group of g equals $\mathrm{Ad}(G)$.

PROOF From Proposition 6.11, we have for the homomorphism $f = \mathrm{Ad}$ that $f(G)$ is a Lie group [a Lie subgroup of $GL(g)$]. Also the Lie algebra of $\mathrm{Ad}(G)$ equals $\mathscr{L}(f(G)) = \mathscr{L}(\mathrm{Im}\, f) = \mathrm{Im}(Tf(e)) = \mathrm{Im}(\mathrm{ad}) = \mathrm{ad}(g)$.

Next if $\phi \in \mathrm{Int}(g)$, then $\phi = \exp D_1 \cdots \exp D_m$ where $D_i = \mathrm{ad}\, X_i$. However, $\exp(\mathrm{ad}\, X_i) = \mathrm{Ad}(\exp X_i)$ so that $\phi = \mathrm{Ad}(\exp X_1 \cdots \exp X_m)$. Since G is connected every element is of the form $\exp X_1 \cdots \exp X_m$ and the results now follow.

Corollary 7.14 Let G be a connected Lie group with Lie algebra g and let $Z(G)$ be the center of G. Then

(a) $Z(G)$ is a Lie group and its Lie algebra is the center of g.

(b) The kernel of the analytic homomorphism $\mathrm{Ad} : G \to \mathrm{Int}(g)$ is $Z(G)$ and $G/Z(G) \cong \mathrm{Int}(g) = \mathrm{Ad}(G)$ as Lie groups. Thus if $Z(G) = \{e\}$, then $\mathrm{Ad} : G \to \mathrm{Int}(g)$ is a Lie group isomorphism.

(c) If the center of g is $\{0\}$, then the center of $\mathrm{Int}(g)$ is $\{I\}$.

PROOF Since $Z(G) = \{b \in G : bxb^{-1} = x \text{ for all } x \in G\}$ we see from connectedness of G that $b \in Z(G)$ if and only if $b(\exp X) = (\exp X)b$ for all $X \in g$. Thus using Corollary 7.12(b)

$$\exp X = b(\exp X)b^{-1} = \exp[\mathrm{Ad}(b)(X)].$$

This is the case if and only if

$$\mathrm{Ad}(b) = I$$

which is the case if and only if $b \in \mathrm{Ker\, Ad}$. Thus we see $Z(G) = \mathrm{Ad}^{-1}(I)$ which is a closed Lie group by Proposition 6.11. Thus $Z(G) = \mathrm{Ker(Ad)}$ so that by Proposition 6.11, the Lie algebra $\mathscr{L}(Z(G)) = \ker(T\,\mathrm{Ad}(e)) = \ker(\mathrm{ad}) = \{Z \in g : \mathrm{ad}\, Z = 0\}$ which is the center of g. The rest of (b) follows from the isomorphism theorem (Corollary 6.16) and (c) is an exercise.

We shall now consider ideals and normal subgroups in more detail and extend the results of Proposition 6.15.

Lemma 7.15 Let G be a Lie group with Lie algebra g, let V be a finite-dimensional vector space over R, and let $\psi : G \to GL(V)$ be an analytic map.

(a) If $\psi : G \to GL(V)$ is a representation of G in V, then $T\psi(e) : g \to gl(V)$ is a representation of g in V.

(b) If ψ is a representation and W is a subspace of V with $\psi(a)W \subset W$ for all $a \in G$, then $[T\psi(e)(X)](W) \subset W$ for all $X \in g$.

(c) Conversely, if G is connected and ψ is a representation and W is a subspace of V with $[T\psi(e)(X)](W) \subset W$ for all $X \in g$, then $\psi(a)W \subset W$ for all $a \in G$.

Proof Part (a) is just a restatement of Proposition 5.19. For (b), let $X \in g$ and $w \in W$. Then since $[T\psi(e)](X)$ is a linear transformation and $\psi(\exp tX) = \exp[tT\psi(e)(X)]$ we have

$$\begin{aligned}[T\psi(e)(X)]w &= [d/dt \exp t(T\psi(e)(X))|_{t=0}]w \\ &= d/dt[(\exp tT\psi(e)(X))w]_{t=0} \\ &= d/dt[\psi(\exp tX)w]_{t=0}.\end{aligned}$$

Thus if for all $a \in G$ we have $\psi(a)w \in W$, then $[T\psi(e)(X)]w \in W$. This is the case because the difference quotient for the derivative and its limit are in W. Conversely for (c) let $[T\psi(e)(X)]w \in W$. Then using the formula $\exp Z = \sum Z^n/n!$ for a linear transformation Z we have $\psi(\exp tX)w = \exp(tT\psi(e)(X))w$ is in W. Thus $\psi(U)w \in W$ for some neighborhood U of e in G and since G is connected and ψ is a homomorphism $\psi(a)w \in W$ for any $a \in G$.

Corollary 7.16 Let G be a connected Lie group with Lie algebra g.

(a) If H is a normal Lie subgroup of G, then its Lie algebra h is an ideal of g.

(b) If h is an ideal of g, then the connected subgroup H generated by $\exp h$ is a normal subgroup of G.

Proof Let $\psi = \mathrm{Ad}$ with $V = g$ in the preceding results. Then for $\phi(a)x = axa^{-1}$ we have $T\phi(a)(e) = \psi(a)$. Thus if H is normal, then for any $a \in G$ we have $\phi(a)H = aHa^{-1} \subset H$ so that, by Corollary 7.12,

$$[T\phi(a)(e)](h) \subset h;$$

that is, $\mathrm{Ad}(a)h \subset h$. Thus for $\psi = \mathrm{Ad}$ we have $\mathrm{ad} = T\psi(e)$ so that by Lemma 7.15 for all $X \in g$,

$$(\mathrm{ad}\ X)(h) = [Xh] \subset h.$$

Thus h is an ideal of g.

Conversely if h is an ideal of g, then $\mathrm{ad}(g)h = [gh] \subset h$. Thus for $\psi = \mathrm{Ad}$ we have $T\psi(e) = \mathrm{ad}$ is such that for any $X \in g$,

$$[T\psi(e)(X)]h = [Xh] \subset h.$$

Therefore by Lemma 7.15(c), $\psi(a)h \subset h$; that is, $\mathrm{Ad}(a)h \subset h$. Consequently,

$$\begin{aligned} a \cdot \exp(h) \cdot a^{-1} &= \phi(a)(\exp h) \\ &= \exp[T\phi(a)(e)(h)] \\ &= \exp[\mathrm{Ad}(a)(h)] \subset \exp h. \end{aligned}$$

Thus since the subgroup H generated by $\exp h$ is connected and $\phi(a)$ is a homomorphism, H is normal.

Exercises (1) If G is a connected Lie group with Lie algebra g, then $\mathrm{Ad}\ G$ is a normal subgroup of $(\mathrm{Aut}\ g)_0$ (the identity component of $\mathrm{Aut}\ g$).

(2) Use Theorem 5.16 to deduce Corollary 7.16(a).

Example (4) Let G be a connected Lie group with Lie algebra g and let H be a closed Lie subgroup with Lie algebra h. The pair (G, H) or (g, h) is called a **reductive pair** if in the Lie algebra g there exists a subspace m such that $g = m + h$ (subspace direct sum) and $(\mathrm{Ad}\ H)(m) \subset m$. In this case the corresponding homogeneous space G/H is called a **reductive homogeneous space**. For the (fixed) decomposition $g = m + h$ we can introduce an algebra multiplication in m as follows: For $X, Y \in m$ let $[XY] = XY + h(X, Y)$ where $XY = [XY]_m$ (respectively $h(X, Y) = [XY]_h$) is the projection of $[XY]$ in g into m (respectively h). Thus m with the multiplication XY becomes an anticommutative algebra; that is, $XY = -YX$. This algebra is analogous to the Lie algebra of a Lie group and can frequently be used to obtain information about the space G/H.

Exercise (3) Show that the Lie algebra identities of g yield the following identities for the above algebra m and the decomposition $g = m + h$.

(i) $XY = -YX$ (bilinear);
(ii) $h(X, Y) = -h(Y, X)$ (bilinear);
(iii) $[Zh(X, Y)] + [Xh(Y, Z)] + [Yh(Z, X)] = (XY)Z + (YZ)X + (ZX)Y$;
(iv) $h(XY, Z) + h(YZ, Y) + h(ZX, Y) = 0$;
(v) $[Ph(X, Y)] = h([PX], Y) + h(X, [PY])$;
(vi) $[P\ XY] = [PX]Y + X[PY]$

for all X, Y, $Z \in m$ and $P \in h$. In particular (vi) says that the mapping $D(P) = \mathrm{ad}_m P : m \to m : X \to [PX]$ is a derivation of the algebra m for each $P \in h$.

(4) Show that Ad $H \subset \mathrm{Aut}(m)$ where m is the above algebra.

As an example consider $G = SO(n)$ and $H = SO(p)$ for $p < n$ as given in Section 6.4. Then relative to the decomposition $g = m + h$ of Section 6.4, one obtains (G, H) or (g, h) as a reductive pair (exercise). Note that the following situation arises: For $p < n - 1$, m is not the zero algebra; that is, $XY \not\equiv 0$. However, for $p = n - 1$ we have m is the zero algebra; that is, $XY \equiv 0$. In general the pair (G, H) or (g, h) is called a **symmetric pair** if there exists a subspace m of g with $g = m + h$ (direct sum) and $\mathrm{Ad}(H)m \subset m$ and $[mm] \subset h$. The corresponding homogeneous space G/H is called a **symmetric space** [Helgason, 1962; Loos, 1969].

Exercise (5) For $n = 2p$ what can be said about the pair (G, H) for $G = GL(n, R)$ and $H = Sp(p, R)$?

CHAPTER 8

SIMPLY CONNECTED LIE GROUPS

In this chapter we review some basic facts on homotopy, fundamental groups, and covering spaces and apply these results to simply connected Lie groups. For example we show that if G is a simply connected Lie group, H is a Lie group, and $f: g \to h$ is a homomorphism of the corresponding Lie algebras, then there is a unique Lie group homomorphism $\psi : G \to H$ with $T\psi(e) = f$. In particular, this implies simply connected Lie groups can be "classified" by their Lie algebras; that is, if G and H are simply connected Lie groups with isomorphic Lie algebras, then G and H are isomorphic Lie groups. This classification is nonvacuous since we show for a given Lie group G with Lie algebra g, there exists a simply connected Lie group $\tilde{G}$ with Lie algebra isomorphic to g. Finally we use various results to show that if G is a connected Lie group, then $\mathrm{Aut}(G)$ is a Lie group.

1. Homotopy Review

In this section we briefly discuss the basics of homotopy, the fundamental group, and show that the fundamental group of a Lie group is commutative.

Definitions 8.1 Let M and N denote (Hausdorff) topological spaces and let I denote the closed interval $[0, 1]$.

(a) Let f_0 and f_1 denote continuous maps $M \to N$. Then f_0 is homotopic to f_1 (denoted by $f_0 \sim f_1$) if there is a continuous map

$$h : M \times I \to N : (x, t) \to h(x, t)$$

satisfying $h(x, 0) = f_0(x)$ and $h(x, 1) = f_1(x)$ for all $x \in M$.

(b) Let $x_0, x_1 \in M$. Then a **path in** M **from** x_0 **to** x_1 is a continuous map (curve) $\alpha : I \to M$ so that $\alpha(0) = x_0$ and $\alpha(1) = x_1$. The **end points** of α are x_0 and x_1 and α is a **closed** path in case $x_0 = x_1$.

(c) Let α and β be paths from x_0 to x_1 in M. Then α and β are **homotopic** or **equivalent** (also denoted by $\alpha \sim \beta$) if there is a continuous map $h : I \times I \to M : (t, s) \to h(t, s)$ satisfying

(i) $h(0, s) = x_0$ and $h(1, s) = x_1$ for all $s \in I$;
(ii) $h(t, 0) = \alpha(t)$ and $h(t, 1) = \beta(t)$ for all $t \in I$.

Thus α and β are homotopic as functions with the additional restriction that the end points are fixed throughout the homotopy. The function h is called a **homotopy** of α to β.

REMARK (1) The homotopy of paths is an equivalence relation. Thus for a path α with endpoints x_0 and x_1 we let $[\alpha]$ denote the equivalence class of α; that is, $[\alpha]$ is the set of all paths homotopic to α.

Definitions 8.2 Let M be a topological space and let $x_0, x_1\, x_2, \in M$.

(a) Let α be a path from x_0 to x_1 and let β be a path from x_1 to x_2. Then the **product** of α and β is the path $\alpha\beta : I \to M$ given by

$$(\alpha\beta)(t) = \begin{cases} \alpha(2t) & \text{for } 0 \le t \le \frac{1}{2} \\ \beta(2t - 1) & \text{for } \frac{1}{2} \le t \le 1. \end{cases}$$

(b) The **inverse** of a path α from x_0 to x_1 is a path α^{-1} from x_1 to x_0 given by

$$\alpha^{-1}(t) = \alpha(1 - t).$$

REMARKS (2) The product of two paths is continuous and so is the inverse.

(3) Let $\alpha_0 \sim \alpha_1$ and $\beta_0 \sim \beta_1$ be paths. Then:

(i) $\alpha_0^{-1} \sim \alpha_1^{-1}$;
(ii) if $\alpha_0\beta_0$ is defined, then $\alpha_1\beta_1$ is defined and $\alpha_0\beta_0 \sim \alpha_1\beta_1$.

This allows us to define the product and inverse of equivalence classes by

$$[\alpha][\beta] = [\alpha\beta] \qquad \text{if } \alpha\beta \text{ is defined}$$

and

$$[\alpha]^{-1} = [\alpha^{-1}].$$

(4) For any z in M let $e_z : I \to M : t \to z$. Then e_z is a path in M with both endpoints equal to z. Let α be a path from x_0 to x_1. Then we have

$$[e_{x_0}][\alpha] = [\alpha] = [\alpha][e_{x_1}].$$

These considerations lead to the following result; see the work of Singer and Thorpe [1967] for various proofs.

Proposition 8.3 Let M be a topological space and let $x_0 \in M$.

(a) The set of equivalence classes of closed paths with endpoints x_0 form a group under the operations of $[\alpha][\beta] = [\alpha\beta]$, $[\alpha]^{-1} = [\alpha^{-1}]$, and identity $[e_{x_0}]$ as above. This group is called the **fundamental group** of M relative to the base point x_0 and denoted by $\pi_1(M, x_0)$.

(b) If M is pathwise connected, then for any $x_0, x_1 \in M$ we have $\pi_1(M, x_0)$ and $\pi_1(M, x_1)$ are isomorphic groups. Thus in this case the fundamental group is essentially independent of the base point and we frequently write $\pi_1(M)$ for $\pi_1(M, x_0)$ and call $\pi_1(M)$ the **fundamental group of** M.

(c) If $f: M \to N$ is a homeomorphism of topological spaces, then the map $f_* : \pi_1(M, x_0) \to \pi_1(N, f(x_0)) : [\alpha] \to [f \circ \alpha]$ is an isomorphism of fundamental groups.

REMARK (5) Let G be a connected Lie group. Then, since it is a manifold, G is also pathwise connected. Thus the fundamental group of G, $\pi_1(G)$, is isomorphic to $\pi_1(G, e)$.

Exercises (1) Let G and H be connected Lie groups. Show $\pi_1(G \times H)$ is isomorphic to $\pi_1(G) \times \pi_1(H)$ as product groups.

(2) Using Theorem 6.20, show that the fundamental group of a connected commutative Lie group is isomorphic to $Z \times \cdots \times Z$, p-times, for some integer p. [Recall that $\pi_1(R) \cong \{0\}$ and $\pi_1(S^1) \cong Z$.]

The following result shows that the fundamental group of a connected Lie group is Abelian.

Proposition 8.4 Let M be a topological space with a continuous "multiplication" function $\mu : M \times M \to M$ such that there exists $e \in M$ with $\mu(x,e) = \mu(e, x) = x$ for all $x \in M$. Then $\pi_1(M, e)$ is Abelian and the product $[\alpha][\beta] = [\mu \circ (\alpha, \beta)]$ in $\pi_1(M, e)$ where $\mu \circ (\alpha, \beta)(t) = \mu(\alpha(t), \beta(t))$.

PROOF Let $[\alpha]$, $[\beta]$ be in $\pi_1(M, e)$ and let $f: I \to M$ be a closed path which is a representative of $[\alpha]$ and similarly let $g : I \to M$ be a representative of $[\beta]$. Since $\mu(e, e) = e$ we can define a closed path $k : I \to M$ with endpoints e by

$$k(t) = \mu(f(t), g(t))$$

for $t \in I$. Then k represents an element $[\gamma]$ in $\pi_1(M, e)$ and $[\gamma]$ depends only on $[\alpha]$ and $[\beta]$. We shall show $[\alpha][\beta] = [\gamma] = [\beta][\alpha]$ and this also shows the formula $[\alpha][\beta] = [\mu \circ (\alpha, \beta)]$.

First to prove $[\gamma] = [\alpha][\beta]$ we can assume that the representative paths f and g satisfy $f(t) = e$ for $\frac{1}{2} \le t$ and $g(t) = e$ for $t \le \frac{1}{2}$. Then using e as a two-sided identity we have

$$k(t) = \begin{cases} f(t) & \text{for} \quad 0 \le t \le \frac{1}{2}, \\ g(t) & \text{for} \quad \frac{1}{2} \le t \le 1. \end{cases}$$

Now we define a continuous map $h : I \times I \to M : (t, s) \to h(t, s)$ by

$$h(t, s) = \begin{cases} f((s+1)t) & \text{for} \quad 0 \le t \le \frac{1}{2}, \quad 0 \le s \le 1, \\ g((t-1)s + t) & \text{for} \quad \frac{1}{2} \le t \le 1, \quad 0 \le s \le 1, \end{cases}$$

and note $h(0, s) = h(1, s) = e$ for all $s \in I$. Also since $h(t, 0) = k(t)$ and since $h(t, 1) = (fg)(t)$, the product fg, we see that k is homotopic to fg; that is, $[\gamma] = [\alpha][\beta]$.

To show $[\gamma] = [\beta][\alpha]$ we first choose the representatives f and g so that $f(t) = e$ for $t \le \frac{1}{2}$ and $g(t) = e$ for $\frac{1}{2} \le t$. Then from $k(t) = \mu(f(t), g(t))$ we see $[\gamma]$ is represented by

$$k(t) = \begin{cases} g(t) & 0 \le t \le \frac{1}{2} \\ f(t) & \frac{1}{2} \le t \le 1. \end{cases}$$

However, as above, k is homotopic to the product gf; that is, $[\gamma] = [\beta][\alpha]$.

Definition 8.5 A topological space M is **simply connected** if it is pathwise connected and its fundamental group $\pi_1(M)$ consists of the identity element. The space M is **locally simply connected** if for each $p \in M$ and each neighborhood U of p there is a simply connected neighborhood V with $p \in V \subset U$.

Theorem 8.6 Let G be a simply connected Lie group with Lie algebra g and let H be a Lie group with Lie algebra h. If $f : g \to h$ is a Lie algebra homomorphism, then there exists a unique Lie group homomorphism $\psi : G \to H$ such that $T\psi(e) = f$.

PROOF Since G is simply connected, it is connected, and consequently by Proposition 5.20 if ψ exists, it is unique. Furthermore from Theorem 6.8 there is a local homomorphism $\phi : G \to H$ with $T\phi(e) = f$. We shall show that we can extend ϕ to the desired homomorphism ψ with $\phi = \psi$ on a suitable nucleus in G. We follow the proof of Hausner and Schwartz [1968].

Thus let U be a connected symmetric nucleus in G so that $\phi(xy) = \phi(x)\phi(y)$ for $x, y \in U$ and also let V be a connected nucleus in G with $V \subset U$ and so that $V^{-1}V \subset U$. For a path $\alpha : I \to G : t \to \alpha(t)$ with endpoints x_0 and x_1 we define a fine partition of I relative to the path α (and U) as follows. Let $0 = t_0 < t_1 < \cdots < t_n = 1$ be a partition of I so that for all subintervals $I_k = [t_{k-1}, t_k]$ for $k = 1, \ldots, n$ we have

$$s, t \in I_k \qquad \text{implies} \qquad \alpha(s)^{-1}\alpha(t) \in U.$$

Such a partition is called **fine**. A fine partition exists as follows. Since I is compact and since the group operations are continuous we have that there exists a $\delta > 0$ so that if $|s - t| < \delta$, then $\alpha(s)^{-1}\alpha(t) \in U$. Also note that a refinement of a fine partition is fine.

For a fixed fine partition as above set $\Delta_k \alpha = \alpha^{-1}(t_{k-1})\alpha(t_k)$ and define

$$F(\alpha) = \phi(\Delta_1 \alpha)\phi(\Delta_2 \alpha) \cdots \phi(\Delta_n \alpha).$$

We now show that $F(\alpha)$ does not depend on the partition of I but depends only on the equivalence class of the path α. First we show that if a point p is added to the above partition, then the value of F is unchanged. Thus we add p to the partition and obtain a new fine partition by considering the interval $[t_{k-1}, t_k]$ being replaced by $[t_{k-1}, p] \cup [p, t_k]$. Then

$$\begin{aligned} \Delta_k \alpha &= \alpha^{-1}(t_{k-1})\alpha(t_k) \\ &= \alpha^{-1}(t_{k-1})\alpha(p)\alpha(p)^{-1}\alpha(t_k) = \Delta'\alpha\Delta''\alpha. \end{aligned}$$

Since $\Delta'\alpha$ and $\Delta''\alpha$ are in U and ϕ is a local homomorphism on U we have

$$\phi(\Delta_k \alpha) = \phi(\Delta'\alpha)\phi(\Delta''\alpha).$$

Thus using this new fine partition we see that the value $F(\alpha)$ remains the same. Therefore since any two fine partitions have a common refinement (which is fine), we see that $F(\alpha)$ does not depend on the partition.

Next $F(\alpha)$ depends only on the equivalence class $[\alpha]$, for let

$$h : I \times I \to G : (t, s) \to h(t, s)$$

be a homotopy with $h(t, 0) = f(t)$ and $h(t, 1) = g(t)$ where both f and g are in $[\alpha]$ (with endpoints x_0 and x_1). Now let $p \in I$ be fixed and let $0 = t_0 < t_1 < \cdots < t_n = 1$ be a fine partition for the path $h(t, p)$. By continuity of h we can choose $s \in I$ sufficiently close to p so that

$$\{t_0, t_1, \ldots, t_n\} \text{ is a fine partition for } h(t, s), \tag{1}$$

$$h(t_k, s)^{-1}h(t_k, p) \in U \qquad \text{for} \quad k = 1, \ldots, n, \tag{2}$$

$$h(t_k, p)^{-1}h(t_k, s) \in U \qquad \text{for} \quad k = 1, \ldots, n. \tag{3}$$

Briefly, for each t_k we can find neighborhood "balls" B_k of p in I so that

$$h(t_{k-1}, a)^{-1}h(t_k, a) \in U$$

for each $a \in B_k$. Thus since there are only finitely many balls B_k, choose the one with the smallest radius B. Then for all $s \in B$ we see $h(t_{k-1}, s)^{-1}h(t_k, s) \in U$ which proves (1). For (2) continue making the finitely many necessary choices of balls to obtain $h(t_k, a)^{-1}h(t_k, p) \in U$ and then take the one with the smallest radius (or B) to obtain (2). For (3) use the fact that U is symmetric ($U^{-1} = U$). We use these equations and induction to obtain for $1 \leq k \leq n$,

$$\begin{aligned}
&\phi[h(t_0, p)^{-1}h(t_1, p)] \cdots \phi[h(t_{k-1}, p)^{-1}h(t_k, p)] \\
&\quad = \phi[h(t_0, s)^{-1}h(t_1, s)] \cdots \phi[h(t_{k-1}, s)^{-1}h(t_k, s)] \cdot \phi[h(t_k, s)^{-1}h(t_k, p)] \qquad (4)
\end{aligned}$$

For $k = 1$ we use ϕ is a local homomorphism on U and $h(t_0, p) = h(t_0, s) = x_0$, and Eqs. (1) and (2) to obtain

$$\begin{aligned}
\phi[h(t_0, p)^{-1}h(t_1, p)] &= \phi[h(t_0, s)^{-1}h(t_1, s) \cdot h(t_1, s)^{-1}h(t_1, p)] \\
&= \phi[h(t_0, s)^{-1}h(t_1, s)]\phi[h(t_1, s)^{-1}h(t_1, p)].
\end{aligned}$$

To pass from k to $k + 1$ we use ϕ as a local homomorphism, Eqs. (1)–(3), and multiply both sides of (4) by

$$\begin{aligned}
&\phi[h(t_k, p)^{-1}h(t_{k+1}, p)] \\
&\quad = \phi[h(t_k, p)^{-1}h(t_k, s) \cdot h(t_k, s)^{-1}h(t_{k+1}, s) \cdot h(t_{k+1}, s)^{-1}h(t_{k+1}, p)] \\
&\quad = \phi[h(t_k, p)^{-1}h(t_k, s)] \cdot \phi[h(t_k, s)^{-1}h(t_{k+1}, s)] \cdot \phi[h(t_{k+1}, s)^{-1}h(t_{k+1}, p)].
\end{aligned}$$

Thus for $k = n$ in (4) we use $h(t_n, p) = h(t_n, s) = x_1$ to obtain

$$F(h(t, p)) = F(h(t, s))$$

for s sufficiently near p in I. Since p is arbitrary in I we can use the transitivity of the homotopy of paths to conclude that if $f, g \in [\alpha]$, then $F(f) = F(g) = F(\alpha)$.

Now let x_0 and x_1 be arbitrary elements in the simply connected group G. Since G is pathwise connected, there exists a path $\alpha : I \to G$ from x_0 to x_1 and any two such curves are homotopic since G is simply connected. Thus we can set

$$F(x_0, x_1) = F(\alpha)$$

because $F(\alpha)$ depends only on the equivalence class $[\alpha]$. In particular for $x_0 = e$ and $x_1 = x$ arbitrary in G we define the desired homomorphism by

$$\psi : G \to H : x \to F(e, x).$$

First we shall show ψ equals ϕ on a connected nucleus V of G where $V \subset U$ and $V^{-1}V \subset U$. For choosing such a neighborhood V we have for $x \in V$ and α a path entirely in V which joins e to x that $0 = t_0 < t_1 = 1$ is a fine partition of I and

$$\psi(x) = \phi[\alpha(t_0)^{-1}\alpha(t_1)] = \phi(\alpha(1)) = \phi(x).$$

Next we prove a few more formulas to show ψ is a homomorphism.

$$F(x, xy) = F(e, y) \qquad \text{for} \quad x, y \in G, \qquad (5)$$

for let α join e to y. Then using $(xa)^{-1}(xb) = a^{-1}b$ we have

$$\begin{aligned}
F(e, y) &= F(\alpha) \\
&= \phi(\Delta_1\alpha)\phi(\Delta_2\alpha)\cdots\phi(\Delta_n\alpha) \\
&= \phi[\alpha(t_0)^{-1}\alpha(t_1)]\cdots\phi[\alpha(t_{n-1})^{-1}\alpha(t_n)] \\
&= \phi[(x\alpha(t_0))^{-1}(x\alpha(t_1))]\cdots\phi[(x\alpha(t_{n-1}))^{-1}(x\alpha(t_n))] \\
&= F(x\alpha) = F(x, xy)
\end{aligned}$$

using the independence of the path for the last equality.

Next let α join e to x and let β join x to xy. Then $F(\alpha\beta) = F(\alpha)F(\beta)$ where $\alpha\beta$ is the product of paths (Definition 8.2). This follows because the interval may be finely partitioned so that the end point of α and the starting point of β are in the partition. Thus since F depends on the equivalence class of a path, $F(\alpha\beta) = F(\alpha)F(\beta)$ gives

$$F(e, xy) = F(e, x)F(x, xy) \qquad \text{for} \quad x, y \in G. \tag{6}$$

Thus for $x, y \in G$ we have

$$\begin{aligned}
\psi(xy) &= F(e, xy) \\
&= F(e, x)F(x, xy), && \text{using (6)} \\
&= F(e, x)F(e, y), && \text{using (5)} \\
&= \psi(x)\psi(y).
\end{aligned}$$

Thus ψ is a homomorphism which is clearly continuous and consequently analytic.

From this result we see that simply connected Lie groups can be "classified" by their Lie algebras as follows.

Corollary 8.7 Let G and H be simply connected Lie groups with Lie algebras g and h. If g and h are isomorphic Lie algebras, then G and H are isomorphic Lie groups.

2. Simply Connected Covering Groups

We shall now show that given a Lie group G with Lie algebra g there exists a simply connected Lie group $\tilde{G}$ with Lie algebra isomorphic to g. As an application, we shall use $\tilde{G}$ to show that $\mathrm{Aut}(G)$ is a Lie group for a connected Lie group G.

Definition 8.8 Let M be a (Hausdorff) topological space. Then:

(a) M is **locally connected** if for each point $p \in M$ and each neighborhood V of p, there exists a connected neighborhood U of p with $U \subset V$;

(b) M is **locally pathwise connected** if for each $p \in M$, every neighborhood of p contains a pathwise connected neighborhood of p.

Exercises (1) Let M be the space in R^2 which is the union of the graph $\sin \pi/x$ for $x \in (0, 1]$ and a path joining the points $(1, 0)$ and $(0, 1)$. Show that M is pathwise connected but not locally pathwise connected.

(2) Is a manifold locally pathwise connected?

Definition 8.9 Let M and $\tilde{M}$ be pathwise connected and locally pathwise connected spaces and let $p : \tilde{M} \to M$ be continuous. Then the pair $(\tilde{M}, p)$ is a **covering space** of M if:

(a) p is surjective;

(b) for each $p \in M$ there exists a neighborhood U of p so that $p^{-1}(U)$ is a disjoint union of open sets, each of which is mapped homeomorphically onto U by p.

REMARK (1) If $(\tilde{M}, p)$ is a covering space of M, then the map $p : \tilde{M} \to M$ is an open map; that is, for each open set $\tilde{U}$ of $\tilde{M}$ we have $p(\tilde{U})$ is open in M [Singer and Thorpe, 1967].

Examples (1) Let $M = S^1$ and $\tilde{M} = R$, and let $p : \tilde{M} \to M : x \to e^{2\pi i x}$. Then $(\tilde{M}, p)$ is a covering space for M. For each $z \in M$ we have that $p^{-1}(z)$ consists of infinitely many points.

(2) Let $M = \tilde{M} = S^1$ and let $p : \tilde{M} \to M : x \to x^2$. Then $(\tilde{M}, p)$ is a covering space of M. In this case $p^{-1}(x)$ consists of two points for each $x \in M$; that is, we have a "double covering."

(3) Let $M = T^2$ $(= S^1 \times S^1)$ and let $\tilde{M} = R^2$. Define $p : \tilde{M} \to M : (x, y) \to (e^{2\pi i x}, e^{2\pi i y})$. Then $(\tilde{M}, p)$ is a covering of M.

We shall assume all spaces are pathwise connected and locally pathwise connected. For the proofs of the following see the work of Chevalley [1946] and Singer and Thorpe [1967].

Theorem 8.10 Let $(\tilde{M}, p)$ be a covering space of M and let N be simply connected.

(a) If $f : N \to M$ is continuous, then there exists a continuous map $\tilde{f} : N \to \tilde{M}$ so that $f = p \circ \tilde{f}$.

(b) If $\tilde{f}, \tilde{g} : N \to \tilde{M}$ are continuous maps so that $p \circ \tilde{f} = p \circ \tilde{g}$ and $\tilde{f}(a) = \tilde{g}(a)$ for some $a \in N$, then $\tilde{f} = \tilde{g}$.

(c) Let $\alpha : I \to M$ be a path in M so that $\alpha(0) = x_0$. Let $\tilde{x}_0 \in \tilde{M}$ be such that $p(\tilde{x}_0) = x_0$. Then there exists a unique path $\tilde{\alpha} : I \to \tilde{M}$ so that $p \circ \tilde{\alpha} = \alpha$ and $\tilde{\alpha}(0) = \tilde{x}_0$.

Theorem 8.11 Let $(\tilde{M}, p)$ be a covering space of M and let $a \in M$, $\tilde{a} \in \tilde{M}$ be such that $p(\tilde{a}) = a$. Then there exists a one-to-one correspondence between $p^{-1}\{a\}$ and the coset space $\pi_1(M, a)/p_* \pi_1(\tilde{M}, \tilde{a})$; recall $p_* \pi_1(\tilde{M}, \tilde{a}) = \{[p \circ \tilde{\alpha}] : [\tilde{\alpha}] \in \pi_1(\tilde{M}, \tilde{a})\}$.

Theorem 8.12 Let M be a pathwise connected, locally pathwise connected, and locally simply connected space. Let H be a subgroup of $\pi_1(M, a)$. Then there is a covering space $(\tilde{M}, p)$ so that $p_* \pi_1(\tilde{M}, \tilde{a}) = H$ where $\tilde{a} \in \tilde{M}$ is such that $p(\tilde{a}) = a$. In case $H = \{e\}$ we see $\tilde{M}$ is simply connected; that is, there is a covering space $(\tilde{M}, p)$ with $\tilde{M}$ simply connected.

PROOF Using the above theorems we sketch the construction of $(\tilde{M}, p)$; for the remaining details, see the proof of Singer and Thorpe [1967]. For motivation, note that if $(\tilde{M}, p)$ exists, then each path $\tilde{\alpha}$ in $\tilde{M}$ starting at $\tilde{\alpha}(0)$ is the unique lift of $\alpha = p \circ \tilde{\alpha}$ in M and also the point $\tilde{\alpha}(1)$ in $\tilde{M}$ is determined by $[\alpha] = [p \circ \tilde{\alpha}]$. Consequently we are led to construct the points of $\tilde{M}$ from paths in M as follows. Let Q be the set of paths in M beginning at the point $a \in M$ and define an equivalence relation $\equiv$ on Q by $\alpha \equiv \beta$ if and only if $\alpha(1) = \beta(1)$ and $[\alpha\beta^{-1}] \in H$. (For the simply connected covering, this is just $\alpha \sim \beta$.) Let $\langle\alpha\rangle$ denote the equivalence class of α under the relation $\equiv$ and let $\tilde{M}$ be the set of all equivalence classes $\langle\alpha\rangle$. We define $p : \tilde{M} \to M$ by $p(\langle\alpha\rangle) = \alpha(1)$ and $(\tilde{M}, p)$ is the desired covering space.

Theorem 8.13 Let G be a connected Lie group. Then there exists a unique simply connected Lie group $\tilde{G}$ which is locally isomorphic to G; that is, G and $\tilde{G}$ have isomorphic Lie algebras. There is a mapping $p : \tilde{G} \to G$ so that $(\tilde{G}, p)$ is a covering space of G and p is a homomorphism and a local isomorphism. Also $\mathrm{Ker}(p)$ is a discrete subgroup of $\tilde{G}$ which is isomorphic to $\pi_1(G)$ and $\mathrm{Ker}(p)$ is in the center of $\tilde{G}$.

PROOF The uniqueness up to isomorphism follows from Corollary 8.7. For the existence, we let $(\tilde{G}, p)$ be the simply connected covering space of G as constructed in Theorem 8.12; the points are the equivalence classes $[\alpha]$ of curves $\alpha : I \to G$ with $\alpha(0) = e$ and $p([\alpha]) = \alpha(1)$. We make $\tilde{G}$ into a group as follows. The product is given by $[\alpha][\beta] = [\gamma]$ where $\gamma(t) = \alpha(t)\beta(t)$ which is well defined since a homotopy of α and a homotopy of β multiply to give a homotopy of γ. The identity is $\tilde{e} = [e]$ where $e(t) = e$ the identity in G. Inverses are given by $[\alpha]^{-1} = [\alpha^{-1}]$ where $\alpha^{-1}(t) = \alpha(t)^{-1}$. The map $p : \tilde{G} \to G$ is a homomorphism because $p([\alpha][\beta]) = \alpha(1)\beta(1) = p([\alpha])p([\beta])$. Next note that p maps a neighborhood of $\tilde{e}$ in $\tilde{G}$ homeomorphically onto a neighborhood of e in G. Consequently the analytic manifold structure on G defines an analytic manifold structure on $\tilde{G}$ with which $\tilde{G}$ becomes a Lie group. Now $p : \tilde{G} \to G$ is a Lie group homomorphism which is a local isomorphism.

Next since p is a local isomorphism $\operatorname{Ker} p$ is a discrete normal subgroup of $\tilde{G}$. Therefore by Proposition 3.25, $\operatorname{Ker} p$ is in the center of $\tilde{G}$. Since $\tilde{G}$ is simply connected, Theorem 8.11 implies that $\operatorname{Ker} p = p^{-1}\{e\}$ is isomorphic to $\pi_1(G)$. Note this argument also shows $\pi_1(G)$ is Abelian.

We now use the simply connected covering group to sketch the proof that the automorphism group of a connected Lie group is a Lie group; for more details see the proofs of Chevalley [1946], Hochschild [1965], and Loos [1969].

Theorem 8.14 Let G be a connected Lie group. Then $\operatorname{Aut}(G)$ is a Lie group.

Proof First let G be simply connected with Lie algebra g. From Theorem 7.1 we know $\operatorname{Aut}(g)$ is a Lie group. Let $\alpha \in \operatorname{Aut}(g)$. Then there is a unique $\theta \in \operatorname{Aut}(G)$ with $(T\theta)(e) = \alpha$ (Theorem 8.6). Using the chain rule, the map $\operatorname{Aut}(G) \to \operatorname{Aut}(g) : \theta \to \alpha$ is a group isomorphism. This isomorphism induces an analytic structure on $\operatorname{Aut}(G)$ where the topology of $\operatorname{Aut}(G)$ is given as follows [Chevalley, 1946; Hochschild, 1965]. Let K be a compact subset of G and let V a neighborhood of e in G. Let $N(K, V)$ denote the set of all elements θ in $\operatorname{Aut}(G)$ so that $\theta(x)x^{-1}$ and $\theta^{-1}(x)x^{-1}$ are in V for every x in K. Then the family of these sets $N(K, V)$ forms a family of nuclei for the identity of $\operatorname{Aut}(G)$. Thus in this case $\operatorname{Aut}(G)$ is a Lie group.

Next suppose G is not simply connected and let $(\tilde{G}, p)$ be the simply connected covering group of G. We shall show that $\operatorname{Aut}(G)$ is isomorphic to a closed subgroup of $\operatorname{Aut}(\tilde{G})$ so that $\operatorname{Aut}(G)$ can be regarded as a Lie group using Theorem 6.9. Now with $N = \tilde{M} = \tilde{G}$ in the notation of Theorem 8.10 we have for $\theta \in \operatorname{Aut}(G)$ that $\theta \circ p : \tilde{G} \to G$ is continuous. Consequently there exists a unique continuous map $\tilde{\theta} : \tilde{G} \to \tilde{G}$ with $\theta \circ p = p \circ \tilde{\theta}$ and $\tilde{\theta}(\tilde{e}) = \tilde{e}$. Let $H = \operatorname{Ker} p$ which is discrete and note that we have the mapping

$$\tilde{G} \times \tilde{G} \to H : (x, y) \to \tilde{\theta}(xy)\tilde{\theta}(y)^{-1}\tilde{\theta}(x)^{-1},$$

for using $\theta \circ p = p \circ \tilde{\theta}$ and p as a homomorphism we see

$$\begin{aligned} p(\tilde{\theta}(xy)\tilde{\theta}(y)^{-1}\tilde{\theta}(x)^{-1}) &= p\tilde{\theta}(xy)p\tilde{\theta}(y)^{-1}p\tilde{\theta}(x)^{-1} \\ &= \theta p(xy)\theta p(y)^{-1}\theta p(x)^{-1} \\ &= \theta(p(x)p(y))\theta p(y)^{-1}\theta p(x)^{-1} = e \end{aligned}$$

using $\theta \in \operatorname{Aut}(G)$. Therefore since $\tilde{G} \times \tilde{G}$ is connected and H is discrete we have from Section 2.3 that $\tilde{\theta}(xy)\tilde{\theta}(y)^{-1}\tilde{\theta}(x)^{-1} = \tilde{e}$ and consequently $\tilde{\theta}$ is an analytic homomorphism of $\tilde{G}$. Since θ is an automorphism we also easily obtain that $\tilde{\theta}$ is an automorphism. Now note for $x \in H = \operatorname{Ker} p$ that $e = \theta(p(x)) = \theta p(x) = p\tilde{\theta}(x)$ so that $\tilde{\theta}(H) \subset H$. Also we have $H \subset \tilde{\theta}(H)$. Thus $\tilde{\theta}(H) = H$. Therefore if we let $K = \{\tilde{\theta} \in \operatorname{Aut}(\tilde{G}) : \tilde{\theta}(H) = H\}$ we obtain K as a subgroup of $\operatorname{Aut}(\tilde{G})$ and a homomorphism $\operatorname{Aut}(G) \to K : \theta \to \tilde{\theta}$. Also K is a Lie subgroup of $\operatorname{Aut}(\tilde{G})$ using H is closed and Theorem 6.9.

Conversely, let $\tilde{\theta} \in K$, let $x \in G$ and let $\tilde{x}$ be any element in $\tilde{G}$ so that $p(\tilde{x}) = x$. Then noting that the value $p(\tilde{\theta}(\tilde{x}))$ depends on x and not on the choice $\tilde{x}$, we set $\theta(x) = p(\tilde{\theta}(\tilde{x}))$. Then θ is an analytic automorphism of G. Thus the mapping $\operatorname{Aut}(G) \to K : \theta \to \tilde{\theta}$ is an isomorphism which makes $\operatorname{Aut}(G)$ into a Lie group.

REMARKS (2) If G_0 denotes the identity component of a Lie group G, it is shown in Loos [1969] that if G/G_0 is finitely generated, then Aut G is a Lie transformation group acting on G.

(3) For the construction of the simply connected covering group $(\tilde{G}, p)$ of specific groups G we refer to the work of Chevalley [1946], Freudenthal and deVries [1969], and Tits [1965].

CHAPTER 9

SOME ALGEBRA

Since a more algebraic approach will be taken in the remaining chapters, we now introduce some of the necessary algebra. Many of the proofs do not actually depend on the use of real or complex numbers but only on the characteristic; consequently all the fields we use will be of characteristic 0. First we discuss tensor products of vector spaces and linear transformations. Using this we consider how to extend the underlying field of the vector space to its algebraic closure and apply this to real Lie algebras. Thus we discuss the complexification, the realification, and real forms of a Lie algebra. Next elementary results on semisimple (i.e. completely reducible) associative algebra and Lie algebra modules are derived and finally Cayley algebras are considered.

1. Tensor Products

Since we shall eventually compute characteristic roots, compare real and complex Lie algebras, etc., we now review some general concepts concerning tensor products.

Definition 9.1 Let V and W be finite-dimensional vector spaces over a field K of characteristic 0. A **tensor product** over K of the vector spaces V and W is a vector space T over K together with a bilinear map

$$\tau : V \times W \to T$$

so that for every bilinear map

$$B : V \times W \to E,$$

where E is any vector space over K, there exists a unique vector space homomorphism $h : T \to E$ satisfying $h \circ \tau = B$; that is, the accompanying diagram is commutative.

$$\begin{tikzcd} V \times W \arrow[rr, "\tau"] \arrow[dr, "B"'] & & T \arrow[dl, "h"] \\ & E & \end{tikzcd}$$

Thus the bilinear map can be factored by the linear map h and the unique "universal" bilinear map τ. The following facts are proved by Jacobson [1953, Vol. II] and Lang [1965].

Theorem 9.2 Let V and W be vector spaces over K.

(a) A tensor product of V and W over K exists.

(b) If (T, τ) and (T', τ') are tensor products over K of V and W, then there exists a unique isomorphism $f : T \to T'$ so that $f \circ \tau = \tau'$. Thus tensor products are unique up to isomorphism and we speak of "the" tensor product.

(c) If $\{X_1, \ldots, X_n\}$ is a basis for V over K and $\{Y_1, \ldots, Y_m\}$ is a basis for W over K, then $\{\tau(X_i, Y_j) : i = 1, \ldots, n \text{ and } j = 1, \ldots, m\}$ is a basis of the tensor product T over K. Thus T is finite-dimensional over K and $\dim T = (\dim V)(\dim W)$.

We shall use the notation

$$V \otimes_K W \qquad \text{or} \qquad V \otimes W$$

for the tensor product T and

$$X \otimes Y$$

for the elements $\tau(X, Y)$ in T. Thus elements in $V \otimes W$ are finite sums $\sum X_i \otimes Y_i$ for $X_i \in V$ and $Y_i \in W$. This uses

$$a(X \otimes Y) = aX \otimes Y = X \otimes aY$$

for $X \in V$, $Y \in W$, and $a \in K$.

We now consider tensor products of homomorphisms. Let $S : V \to V'$ and $T : W \to W'$ be homomorphisms of vector spaces over K and let

$$S \times T : V \times W \to V' \times W' : (X, Y) \to (S(X), T(Y)).$$

For the tensor products $(V \otimes W, \tau)$ and $(V' \otimes W', \tau')$ we note that

$$\tau' \circ (S \times T) : V \times W \to V' \otimes W'$$

is bilinear. Thus by definition, there exists a unique vector space homomorphism $P : V \otimes W \to V' \otimes W'$ so that $P \circ \tau = \tau' \circ (S \times T)$. From this we obtain

$$P(X \otimes Y) = S(X) \otimes T(Y)$$

and we use the notation $P = S \otimes T$ or $S \otimes_K T$ called the **tensor product of the homomorphisms** S and T.

Now let A and B be nonassociative algebras over the field K. We shall now construct a multiplication on the tensor product $T = A \otimes_K B$ as follows. Since T is generated by elements of the form $X \otimes Y$, we define for $X = \sum X_i \otimes Y_i$ and $X' = \sum X_j' \otimes Y_j'$ in T the bilinear function

$$\mu : T \times T \to T : (X, X') \to \sum X_i X_j' \otimes Y_i Y_j'.$$

Thus $A \otimes_K B$ becomes an algebra over K called the **tensor product of the algebras** A and B.

Exercises (1) How can an algebra A with bilinear multiplication $\mu : A \times A \to A$ be defined in terms of $A \otimes A$?

(2) Let U, V, W be vector spaces over K. Then show the following isomorphisms

$$V \otimes_K K \cong V \cong K \otimes_K V, \qquad V \otimes_K W \cong W \otimes_K V,$$
$$U \otimes_K (V \otimes_K W) \cong (U \otimes_K V) \otimes_K W,$$
$$\mathrm{End}_K(V) \otimes \mathrm{End}_K(W) \cong \mathrm{End}_K(V \otimes W),$$
$$V = \sum V_i \quad \text{and} \quad W = \sum W_j \qquad \text{implies} \qquad V \otimes W \cong \sum V_i \otimes W_j.$$

(3) If V and W are nonzero vector spaces over K, show that the tensor map $\tau : V \times W \to V \otimes W$ is not injective.

(4) Let $S : V \to V'$ and $T : W \to W'$ be vector space homomorphisms. Find the kernel of $S \otimes T$. Show that if S and T are isomorphisms, so is $S \otimes T$.

(5) Let A and B be associative algebras over K. Show that $A \otimes_K B$ is an associative algebra over K.

2. Extension of the Base Field

We continue the notation of the preceding section and discuss how one can extend the base field of a vector space to a larger field—in particular, the extension to the algebraic closure of the original field; see the work of Jacobson [1953, Vol. II; 1962] and Lang [1965]. We also consider extensions of homomorphisms and discuss in detail the real and complex case in the next section.

Definition 9.3 Let V be a vector space over the field K and let P be a field extension of K which is a finite-dimensional vector space over K. Form the tensor product $P \otimes_K V$ and regard it as a vector space over P by defining

$$p(\sum p_i \otimes X_i) = \sum pp_i \otimes X_i.$$

This is well defined and the vector space axioms for $P \otimes_K V$ over P hold. We denote this vector space over P by $V(P)$ and call it the vector space obtained from V by **extending the base field** K to the field P.

Proposition 9.4 Let V be a vector space over K with basis $X_1, \ldots, X_m$. Then the vectors $1 \otimes X_1, \ldots, 1 \otimes X_m$ form a basis of $V(P)$ over P where P is a field extension of K. Thus the dimension of V over K equals the dimension of $V(P)$ over P.

PROOF Let $\bar{X}_i = 1 \otimes X_i$. Then for $p \in P$ we have $p\bar{X}_i = p(1 \otimes X_i) = p \otimes X_i$. Thus since any X in $V(P)$ has the form $\sum p_i \otimes X_i$, X also has the form $\sum p_i \bar{X}_i$. Thus the $\bar{X}_i$ are generators for $V(P)$ over P and they are also linearly independent over P (exercise).

REMARKS (1) The vector space $V(P) = P \otimes_K V$ is also a vector space over K and in this case $\dim_K V(P) = (\dim_K P)(\dim_K V)$ as noted in Theorem 9.2.

(2) For a basis $X_1, \ldots, X_m$ of V over K, the set of K-linear combinations of the elements $1 \otimes X_1, \ldots, 1 \otimes X_m$ form a subset $\bar{V} = \{1 \otimes X : X \in V\}$ of $V(P)$. Then $\bar{V}$ is a K-subspace of $V(P)$ and the map $V \to \bar{V} : X \to 1 \otimes X$ is a K-vector space isomorphism. Thus V can be identified as a K-subspace of $V(P)$ and $\bar{V}$ satisfies the following:

(i) The vector space spanned by $\bar{V}$ over P equals $V(P)$.

(ii) If $\bar{N}$ is a subset of $\bar{V}$ consisting of linearly independent vectors over K, then $\bar{N}$ consists of linearly independent vectors over P.

Next let A be a nonassociative algebra over K with P a field extension of K and let $A(P) = P \otimes_K A$ be the vector space obtained by extending the base field. Then regarding $A(P)$ as the tensor algebra of P and A as in Section 9.1, $A(P)$ becomes an algebra over P by the multiplication

$$(\sum p_i \otimes X_i)(\sum q_j \otimes Y_j) = \sum p_i q_j \otimes X_i Y_j.$$

(3) Let $X_1, \ldots, X_m$ be a basis of the algebra A over K and let the structure constants $c_{ij}^k \in K$ be given by

$$X_i X_j = \sum_k c_{ij}^k X_k.$$

As above the vectors $\overline{X}_1, \ldots, \overline{X}_n$ with $\overline{X}_i = 1 \otimes X_i$ are a basis of $A(P)$ over P, and from various definitions

$$\begin{aligned}\overline{X}_i \overline{X}_j &= (1 \otimes X_i)(1 \otimes X_j) \\ &= 1 \otimes X_i X_j = \sum_k c_{ij}^k \overline{X}_k;\end{aligned}$$

that is, the structure constants relative to corresponding basis in $A(P)$ are the same.

Exercise (1) Let A be an associative or Lie algebra over K and let P be a field extension of K. Show that the corresponding extension $A(P)$ is an associative or Lie algebra.

We next consider the extension of the linear transformation $T\colon V \to W$ of vector spaces over K. Thus since $V(P) = P \otimes_K V$ and $W(P) = P \otimes_K W$ we let

$$\overline{T} = I \otimes T : V(P) \to W(P)$$

be as in Section 9.1. Then $\overline{T}$ is called the **extension** of T and is specifically given by

$$\overline{T}(\sum p_i \otimes X_i) = \sum p_i \otimes TX_i.$$

In particular, if $X_1, \ldots, X_n$ is a basis of V and $T(X_i) = \sum a_{ji} X_j$, then for the corresponding basis $\overline{X}_1, \ldots, \overline{X}_n$ of $V(P)$ we have

$$\overline{T}(\overline{X}_i) = 1 \otimes TX_i = \sum a_{ji} \overline{X}_j.$$

Thus the matrix of $\overline{T}$ relative to $\overline{X}_1, \ldots, \overline{X}_n$ is the same as the matrix of T relative to $X_1, \ldots, X_n$.

REMARK (4) A variation of these results is frequently applied when the extension field P is the algebraic closure of K; in particular when the base field R is the real numbers and the extension C is the complex numbers. Thus one starts with a real vector space V and an endomorphism $T\colon V \to V$ for which one needs to know information about the characteristic roots. Then we pass to the extension $\overline{T}\colon V(C) \to V(C)$ to compute this information. Frequently the results are already in R or one can prove certain results hold over R if and only if they hold over C.

More specifically, let V be a vector space over K. Then an endomorphism $T: V \to V$ is called **split** if all the characteristic roots of T are in K. An associative or Lie algebra A over K is called **split** if all the left multiplications $L(X): A \to A : Y \to XY$ and all the right multiplications $R(X): A \to A : Y \to YX$ are split endomorphisms.

3. Complexification

We now apply the preceding results to the case when the base field is the real numbers R and the extension field is the complex numbers C. We develop some terminology for this case and give an example of a real simple Lie algebra for which the complex extension is not simple.

Let V be a vector space over R. Then noting that for $X, Y \in V$ we have $X + i \otimes Y \in V(C) = C \otimes_R V$, we can formally think of

$$V(C) = \{X + iY : X, Y \in V \quad \text{and} \quad i = (-1)^{1/2}\}$$

with the complex number multiplication

$$(a + ib)(X + iY) = (aX - bY) + i(bX + aY).$$

Note that $V \subset V(C)$ by identifying $V = V + i\{0\}$.

Definition 9.5 Let g be a Lie algebra over R. Then the **complexification** of g is the Lie algebra $g(C) = C \otimes_R g$.

We also use the notation

$$\tilde{g} = g + ig$$

and note that the multiplication in $\tilde{g}$ is given by

$$[U + iV \quad X + iY] = [UX] - [VY] + i([VX] + [UY]).$$

Example (1) We now consider various ways of obtaining complex and real Lie algebras from a given Lie algebra. Thus let g be the three-dimensional Lie algebra with basis E, F, H over C with multiplication given by

$$[HE] = 2E, \qquad [HF] = -2F, \qquad [EF] = H.$$

Then g is a simple Lie algebra; that is, g has no proper ideals (Definition 6.10). For suppose h is a subspace of g so that $[gh] \subset h$ and let

$$X = aH + bE + cF \in h$$

with $a \neq 0$. Then from the multiplicative relations

$$[X \quad E] = 2aE - cH \in h$$

so that $4aE = [H[XE]]$ is in h. Thus since $a \neq 0$ we see E is in h. From the multiplicative relations of g this implies $h = g$. If $a = 0$, then similar arguments also show $h = g$.

Next let $X = iH$, $Y = iE$, and $Z = iF$ be in g and let g_R be the six-dimensional vector space over R with basis H, E, F, X, Y, Z. With the multiplication in g_R induced from g; for example,

$$[XY] = [iH\, iE] = i^2[HE] = -2E,$$

we obtain the accompanying multiplication table for g_R, where $*$ is computed

	H	E	F	X	Y	Z
H	0	$2E$	$-2F$	0	$2Y$	$-2Z$
E		0	H	$-2Y$	0	X
F			0	$2Z$	$-X$	0
X				0	$-2E$	$2F$
Y		$*$			0	$-H$
Z						0

using the anticommutivity of the multiplication. Thus g_R becomes a six-dimensional Lie algebra over R which is simple (exercise).

Now let $\tilde{g}_R = C \otimes_R g_R$ be the complexification of g_R as above. Then $\tilde{g}_R$ is six-dimensional over C but g is three-dimensional over C. With the basis H, E, F, X, Y, Z of $\tilde{g}_R$ over C let k (respectively $\bar{k}$) be the complex subspace of $\tilde{g}_R$ spanned by

$$H + iX, \quad E + iY, \quad F + iZ$$

(respectively $H - iX$, $E - iY$, $F - iZ$). From the multiplicative relations in the above table we obtain $[kk] = k$, $[k\bar{k}] = \{0\}$, and $[\bar{k}\bar{k}] = \bar{k}$ so that k and $\bar{k}$ are ideals of $\tilde{g}_R$. Also note the direct sum

$$\tilde{g}_R = k \oplus \bar{k}.$$

The C-linear map $\phi : g \to k$ given by

$$\phi(H) = \tfrac{1}{2}(H + iX), \qquad \phi(E) = \tfrac{1}{2}(E + iY), \qquad \phi(F) = \tfrac{1}{2}(F + iZ)$$

is a Lie algebra isomorphism. Similarly $\bar{k}$ is isomorphic to g. In summary, we see that the simple complex Lie algebra g yields a simple real Lie algebra g_R. However, the complexification $\tilde{g}_R$ of this real simple algebra is not a simple complex Lie algebra.

Example (2) There exist real Lie algebras which are not isomorphic but have the same complexification. Let $n > 1$ and let $0 \le k \le [n/2]$, where $[n/2]$ means the largest integer less than or equal to $n/2$. Let

$$\sigma_k = \begin{bmatrix} -I_k & 0 \\ 0 & I_{n-k} \end{bmatrix}$$

where I_j is the $j \times j$ identity matrix. Then σ_k defines a symmetric nondegenerate bilinear form B_k on the real vector space $V = R^n$ [note example (7), Section 2.3]. Let

$$SO(k, n-k) = \{T \in SL(R^n) : B_k(TX, TY) = B_k(X, Y) \quad \text{all} \quad X, Y \in R^n\}$$

which has Lie algebra

$$so(k, n-k) = \{S \in sl(R^n) : B_k(SX, Y) + B_k(X, SY) = 0 \quad \text{all} \quad X, Y \in R^n\}.$$

Then $so(k, n-k)$ and $so(1, n-1)$ are not isomorphic if $k \ne 1$ (exercise).
However, the complexification of $so(k, n-k)$ for any k is just

$$so(n, C) = \{S \in gl(C^n) : \; B(SX, Y) + B(X, SY) = 0 \quad \text{all} \quad X, Y \in C^n\}$$

where B is the complex symmetric bilinear form with matrix $\sigma = I_n$.

Exercise (1) A **conjugation** in a complex Lie algebra g is a function $C : g \to g$ so that for $X, Y \in g$, $a \in C$, and $\bar{a}$ the conjugate of a in C, we have

$$C[XY] = [CX \quad CY], \qquad C^2 = I,$$
$$C(X+Y) = CX + CY, \qquad C(aX) = \bar{a}C(X).$$

(i) Let g be a real Lie algebra and let $\tilde{g} = g + ig$ be its complexification. Show that $C : \tilde{g} \to \tilde{g} : X + iY \to X - iY$ is a conjugation and g is the fixed point set of C.

(ii) For the Lie algebras k and $\bar{k}$ in the above example (1), show that there exists a conjugation of g such that $C : k \to \bar{k}$.

(iii) Let h be a real Lie subalgebra of the Lie algebra g and let $\tilde{g}$ be the complexification of g as in (i). Show that the complexification $\tilde{h} = C \otimes_R h$ is a subalgebra of $\tilde{g}$. Conversely, show that if $\tilde{h}$ is a subalgebra of $\tilde{g}$ so that $C(\tilde{h}) \subset \tilde{h}$ where C is given in (i), then $\tilde{h}$ is the complexification of some real subalgebra of g.

Definition 9.6 Let g be a Lie algebra over C of complex dimension n.

(a) By restricting the scalars to R, the Lie algebra g can be considered as a Lie algebra of dimension $2n$ over R denoted by g_R and is called the **realization** or **realification** of g.

(b) Let g_R be the realization of the complex Lie algebra g, let h be a real subalgebra of g_R, and let $\tilde{h} = h + ih$ be the complexification of h. Then h is called a **real form** of g provided there exists a complex Lie algebra isomorphism $\phi : g \to \tilde{h}$ so that $\phi(X) = X$ for all $X \in h$.

REMARK (1) Frequently in the definition of real form, it is required that $g = \tilde{h}$.

Example (3) Let g be the three-dimensional complex Lie algebra of example (1), this section. Then the six-dimensional algebra g_R is the realization of g. The real subalgebra h of g_R generated by $\{H, E, F\}$ is a real form of g.

Exercise (2) Is the realization k_R of k in example (1) a real form of $\tilde{g}_R$?

Proposition 9.7 Let g be a complex Lie algebra and let h be a subset of g. Then h is a real form of g if and only if h is the set of fixed points of a conjugation of g.

PROOF Let C be a conjugation of g which has h as its set of fixed points. Then clearly h is a subalgebra of the realification g_R. Next let $\tilde{h}$ be the complexification of h and define a map

$$\phi : g \to \tilde{h} : X \to \frac{X + C(X)}{2} + i\left(\frac{X - C(X)}{2i}\right)$$

which is an isomorphism so that for all $X \in h$

$$\phi(X) = C(X) = X.$$

Thus h is a real form of g.

Conversely if h is a real form of g, we can identify g with $\tilde{h}$ by the isomorphism in the definition of real form; that is, let $g = h + ih$ as a direct sum. Then the map

$$C : g \to g : X + iY \to X - iY$$

is a conjugation with h as its set of fixed points.

Exercises (3) Show that the map ϕ defined in the above proof actually maps g into the complexification of h. Thus what can be said about the expressions $(X + C(X))/2$ and $(X - C(X))/2i$ for all $X \in g$?

(4) Let h be a real form of $g = h + ih$, and let $C : g \to g : X + iY \to X - iY$ be a conjugation. Let $\tilde{C}$ be another conjugation of g with h as its fixed point set. Show that $C = \tilde{C}$. What can be said if we regard $g \cong h + ih$?

(5) Show that there exists a complex Lie algebra which has no real form, possibly as follows:

(i) Let $g = CX + CY + CZ$ where the basis, X, Y, Z of this three-dimensional algebra has multiplication $[XY] = \alpha Y$, $[XZ] = \beta Z$, $[YZ] = 0$ for some α, $\beta \in C$. Then $g' = [gg] = CY + CZ$.

(ii) Assume h is a real form of g. Then h is the set of fixed points of a conjugation C of g. Show that $h' = g' \cap h$ is a real form for g'.

(iii) Let $U = aX + bY + cZ$ be in h but not in h' so that $a \neq 0$. Then $\mathrm{ad}_h U$ leaves h' invariant and so induces a real endomorphism of h'. Show that $(\mathrm{ad}\, U)Y = \alpha a Y$ and $(\mathrm{ad}\, U)Z = \beta a Z$ so that αa and βa are real characteristic roots. Thus $\alpha/\beta = \alpha a/\beta a \in R$. However α and β can be chosen in C so this cannot happen which contradicts the assumption of a real form.

Definition 9.8 Let g be a complex Lie algebra. The **conjugate** Lie algebra g^* of g is given as follows. The algebra g^* is the same Abelian group as g and has the same algebra multiplication as g. However, the scalar multiplication $*$ in g^* is given by $a * X = \bar{a}X$, where $\bar{a}$ is the complex conjugate of $a \in C$ and the scalar multiplication $\bar{a}X$ is that in g.

With these definitions g^* is clearly a Lie algebra over C.

Theorem 9.9 Let g be a complex Lie algebra and let g_R be its realization and $\tilde{g}_R$ be the complexification of g_R. Then

$$\tilde{g}_R \cong g \oplus g^*$$

where the (external) direct sum of g and its conjugate algebra g^* contains these algebras as ideals.

PROOF If $A \dotplus B$ denotes the elements in $g \oplus g^*$, then we define the map

$$T : \tilde{g}_R \to g \oplus g^* : X + iY \to (X + iY) \dotplus (X + i * Y)$$

where the X, $Y \in g_R$ are uniquely determined by $X + iY \in \tilde{g}_R = C \otimes g_R$. Thus the elements $X + iY \in g$ and $X + i * Y \in g^*$ are well defined as elements of these algebras over C. Therefore T is well defined and is a Lie algebra isomorphism. For example, let $X + iY$, $U + iV \in \tilde{g}_R$. Then

$$[X + iY \quad U + iV[= [XU] - [YV] + i([YU] + [XV])$$

in $\tilde{g}_R$ gives

$$\begin{aligned}T([X+iY \quad U+iV]) &= ([XU]-[YV]+i([YU]+[XV])) \\ &\quad \dotplus ([XU]-[YV]+i*([YU]+[XV])) \\ &= ([XU]-[YV]+[iYU]+[XiV)]) \\ &\quad \dotplus ([XU]-[YV]+[i*YU]+[Xi*V]) \\ &= [X+iY \; U+iV] \dotplus [X+i*Y \; U+i*V].\end{aligned}$$

However from the definition of multiplication in a direct sum we also have

$$\begin{aligned}[T(X+iY)T(U+iV)] &= [(X+iY)\dotplus(X+i*Y)\,(U+iV)\dotplus(U+i*V)] \\ &= [X+iY \; U+iV] \dotplus [X+i*Y \; U+i*V]\end{aligned}$$

which shows T preserves products. The proof that T is a vector space isomorphism is left to the reader.

Exercise (6) Compare example (1) of this section with Theorem 9.9.

4. Modules and Representations

We briefly review the basics of modules and representations for associative algebras and make the corresponding definitions for Lie algebras. However, throughout most of the text we shall view these concepts in the framework of algebras of endomorphisms acting on a vector space.

Definition 9.10 Let A be an associative algebra over a field K and let V be a vector space over K. Then V is a (left) A-**module** provided there exists a bilinear mapping $A \times V \to V : (S, X) \to SX$ satisfying $(ST)X = S(TX)$ for all $S, T \in A$, and $X \in V$, and $a(SX) = (aS)X = S(aX)$ for all $a \in K$.

REMARKS (1) Given an A-module V, for each $S \in A$ we can define the endomorphism $\rho(S) : V \to V$ by $\rho(S)X = SX$. Thus since $\rho(S)\rho(T) = \rho(ST)$ we see that the action of A on V is given by the action of the algebra of endomorphisms $\rho(A) = \{\rho(S) : S \in A\}$ on V. Conversely if A is an associative algebra acting on V according to the above formulas, then V is an A-module.

(2) According to Definition 7.8, the above mapping $\rho : A \to \rho(A) : S \to \rho(S)$ is a representation of A in V. Conversely given a representation ρ of A in V, then we can make V into an A-module by defining the action $A \times V \to V : (S, X) \to \rho(S)X$. Thus the concepts of an associative algebra of

endomorphisms acting on a vector space, A-modules, and representations are the same.

(3) We assume the reader is familiar with submodules, quotient modules, module homomorphisms, etc., but we discuss completely reducible (i.e., semisimple) modules in the next section.

Definition 9.11 Let g be a Lie algebra over a field K and let V be a vector space over K. Then V is a g-**module** provided there exists a bilinear mapping $g \times V \to V : (S, X) \to SX$ satisfying

$$[ST]X = S(TX) - T(SX)$$

for all $S, T \in g$ and $X \in V$, and $a(SX) = (aS)X = S(aX)$ for all $a \in K$.

Remarks (4) As for the associative case, one has the definitions of Lie submodules, quotient modules, homomorphisms, etc. Thus if V and W are g-modules, then $\phi : V \to W$ is a g-homomorphism if ϕ is a homomorphism of the vector spaces and commutes with the action of g; that is, $\phi(SX) = S(\phi X)$ for all $S \in g$ and $X \in V$.

(5) If V is a g-module, then for each $S \in g$ we can define the endomorphism $\rho(S) : V \to V$ by $\rho(S)X = SX$. From the definition we see

$$\rho([ST]) = \rho(S)\rho(T) - \rho(T)\rho(S) = [\rho(S), \rho(T)]$$

so that the mapping $\rho : g \to \rho(g) : S \to \rho(S)$ is a representation of g in V according to Definition 7.8. Thus a g-module yields a representation and also conversely; for if ρ is a representation of g in V, then V becomes a g-module by defining $g \times V \to V : (S, X) \to \rho(S)X$.

(6) Let P be a set of linear transformations acting on the finite-dimensional vector space V over K; that is, we have a mapping $P \times V \to V : (S, X) \to SX$ so that $S(aX + bY) = aS(X) + bS(Y)$ and we can regard $P \subset \mathrm{End}(V)$. Then we can form the associative algebra of endomorphisms $\mathscr{A}(P)$ or the Lie algebra of endomorphisms $\mathscr{L}(P)$ generated by P. Thus $\mathscr{A}(P)$ consists of all finite sums of products of elements of P, and $\mathscr{L}(P)$ consists of all finite sums of commutators of elements of P; note Proposition 7.6. In either case we can regard V as an $\mathscr{A}(P)$-module or a $\mathscr{L}(P)$-module, and the P-invariant subspaces are just submodules.

Examples (1) Let V be a vector space over R, let G be a Lie group, and let $\rho : G \to GL(V)$ be a (differentiable) representation of G in V. Thus regarding V as a manifold, G operates differentiably on V by the action

$$G \times V \to V : (S, X) \to \rho(S)X$$

(recall Definition 3.17). From Lemma 7.15 we see that if g is the Lie algebra of G, then $T\rho(e) : g \to gl(V)$ is a representation of g in V. Thus V is a g-module.

A more direct computational way of viewing this is to regard $G \subset GL(V)$ with Lie algebra $g \subset gl(V)$. Then for $S, T \in g$ we have for t near 0 in R that for $X \in V$,

$$(e^{tS}e^{tT})X = e^{tS}(e^{tT}X).$$

Using this formula and the Campbell–Hausdorff formula for computing $\exp tS \exp tT = \exp(tS + tT + \frac{1}{2}t^2[S, T] + \cdots)$ one obtains

$$[S, T]X = S(TX) - T(SX)$$

as expected.

(2) We can construct more modules using tensor products as follows. First let A be an associative algebra over the field K and let V and W be A-modules. Then $V \otimes_K W$ becomes an A-module when the action of A is given by $P(\sum X_i \otimes Y_i) = \sum PX_i \otimes PY_i$ for $P \in A$ and $X_i \in V$, $Y_i \in W$. Similarly if g is a Lie algebra over K and V and W are g-modules, then $V \otimes_K W$ becomes a g-module when the action of g is given by

$$P(\sum X_i \otimes Y_i) = \sum PX_i \otimes Y_i + X_i \otimes Py_i$$

for $P \in g$ and $X_i \in V$, $Y_i \in W$ (that is, "differentiate" the associative action).

5. Semisimple Modules

We continue the notation of the preceding section and discuss irreducible modules and their direct sum. We also review the basics of finite-dimensional semisimple associative algebras and their modules; for elementary references see the work of Jacobson [1953, Vol. II; 1962], Lang [1965], and Paley and Weichsel [1966].

Definition 9.12 Let V be a finite-dimensional nonzero vector space over K. Let A be an associative subalgebra of $\mathrm{End}(V)$ or let A be a Lie subalgebra of $gl(V)$, thus V is either an associative or Lie module.

(a) The vector space V is a **simple** or **irreducible** A-module if the only A-submodules of V are V and $\{0\}$. In this case A is called an **irreducible algebra** of endomorphisms on V and say A acts irreducibly on V.

(b) The vector space V is a **semisimple** or **completely reducible** A-module if V is a vector space direct sum of irreducible A-modules. In this case we frequently say that A acts in a completely reducible manner on V or A is a **completely reducible algebra** of endomorphisms on V.

We summarize some standard results [Jacobson 1953, Vol. II; Lang, 1965].

Proposition 9.13 Let V be a vector space over K and let A be an associative or Lie algebra of endomorphisms of V so that V is an A-module.

(a) Then V is a completely reducible A-module if and only if for every A-submodule W of V there exists an A-submodule W' so that $V = W + W'$ which is a submodule direct sum.

(b) If V is completely reducible with decomposition $V = V_1 + \cdots + V_t$ into irreducible subspaces V_i, then the V_i are uniquely determined up to an A-isomorphism and the length t is unique. In this case, we can choose a basis of V consisting of bases of the components V_i so that each $S \in A$ has block matrix

$$\begin{bmatrix} S_1 & & & 0 \\ & S_2 & & \\ & & \ddots & \\ 0 & & & S_t \end{bmatrix}.$$

The matrix S_i represents the action of S on V_i.

Now we restrict A to be associative.

Proposition 9.14 Let V be a vector space over K and let A be an associative algebra of endomorphisms of V so that V is an A-module. Let $\Delta = \{T \in \mathrm{End}(V) : TS = ST \text{ for all } S \in A\}$ be the centralizer of A in $\mathrm{End}(V)$.

(a) (Schur's lemma) If V is an irreducible A-module, then Δ is a division ring; that is, Δ is an associative algebra over K for which every nonzero element has an inverse. In this case, if the field K is algebraically closed, then $\Delta = KI$ where I is the identity endomorphism.

(b) (Burnside's theorem) If A is an irreducible algebra of endomorphisms on V and if the field K is algebraically closed, then $A = \mathrm{End}_K(V)$.

Examples (1) Let A be a nonassociative algebra over K with bilinear multiplication function α. Then we have previously considered an ideal B of A as a subspace so that $\alpha(B, A) \subset B$ and $\alpha(A, B) \subset B$. Now, as in Section 7.2, we let $R(X)$ and $L(X)$ be the right and left multiplication functions on A and let P be the subspace of $\mathrm{End}(A)$ spanned by all $R(X)$ and $L(Y)$ for $X, Y \in A$. Then an ideal is just a subspace B which is invariant under the associative algebra $\mathscr{A}(P)$ or the Lie algebra $\mathscr{L}(P)$. In particular an algebra A is **simple** if and only if $A^2 \neq \{0\}$ and A has no proper ideals; that is, $A^2 \neq \{0\}$ and A is $\mathscr{A}(P)$- or $\mathscr{L}(P)$-irreducible.

In particular, if A is an associative algebra over K which is simple, then A is isomorphic to a suitable $n \times n$ matrix ring E over some division ring $D \supset K$; that is, the matrix ring of all $n \times n$ matrices (a_{ij}) where $a_{ij} \in D$. See the elementary proofs of Lang [1965] and Paley and Weichsel [1966].

(2) There are many starting points for the definition of a **semisimple** nonassociative algebra—depending on the class of algebras being studied. However, regardless of the starting definition, the usual conclusion is that A is a semisimple algebra if $A^2 \neq \{0\}$ and $A = A_1 \oplus \cdots \oplus A_t$ a direct sum of ideals which are simple algebras. Thus A is a completely reducible $\mathscr{A}(P)$- or $\mathscr{L}(P)$-module.

In particular, let A be a finite-dimensional associative algebra over K. An ideal N of A is nilpotent if there exists an integer k so that $\{0\} = N^k$ $(= NN \ldots N$, k-times). One can show that the sum of two nilpotent ideals is a nilpotent ideal and consequently define the **radical** of A, $\mathrm{rad}(A)$, to be the maximal nilpotent ideal. We shall show in Chapter 12 that if A is an associative algebra so that $A^2 \neq \{0\}$, then $A = A_1 \oplus \cdots \oplus A_t$ is a direct sum of ideals which are simple algebras if and only if $\mathrm{rad}(A) = \{0\}$; that is, A is a semisimple A-module if and only if A has no nonzero nilpotent ideals.

Note that if A is semisimple, then $A = A_1 \oplus \cdots \oplus A_t$ where each A_i is isomorphic to some ring of all $n_i \times n_i$ matrices over some division ring. Thus, in particular, each A_i has an identity e_i so A has an identity $1 = e_1 + \cdots + e_t$.

These remarks can be used to prove the following results.

Proposition 9.15 Let V be a finite-dimensional vector space over K and let A be an associative subalgebra of $\mathrm{End}(V)$ such that V is a completely reducible A-module. Then A is semisimple.

Proof Let $N = \mathrm{rad}(A)$ be the maximal nilpotent ideal in A and let $V = V_1 \oplus \cdots \oplus V_t$ where the V_i are nonzero A-irreducible submodules. Let

$$NV_i = \{\textstyle\sum T_k X_k : T_k \in N \quad \text{and} \quad X_k \in V_i\}.$$

Then since N is an ideal in A we see that NV_i is an A-submodule of V which is contained in V_i. Since V_i is irreducible, NV_i equals V_i or $\{0\}$. If $NV_i = V_i$, then

$$N^2 V_i = N(NV_i) = NV_i = V_i$$

and by induction $V_i = N^k V_i$. However, since N is nilpotent, this equation implies $V_i = \{0\}$, a contradiction. Thus $NV_i = \{0\}$ for all i, so that $N = \{0\}$; that is, A is semisimple.

Remarks (1) Before proving the converse statement we briefly review results on a simple associative algebra A over K in terms of an $n \times n$ matrix algebra E over a division algebra $D \supset K$; see the more formal proofs of Lang [1965] and Paley and Weichsel [1966]. First the identity matrix $I = E_1 + \cdots + E_n$ where $E_i^2 = E_i$ are idempotents which are matrices of the

$$E_i = \begin{bmatrix} 0 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{bmatrix}$$

form with 1 in the (i, i)-position. Thus EE_i is an irreducible E-submodule consisting of matrices of the form

$$\begin{bmatrix} & a_{1i} & \\ 0 & \vdots & 0 \\ & a_{ni} & \end{bmatrix}$$

where $a_{pi} \in D$. Consequently $E = EE_1 + \cdots + EE_n$ is the direct sum of these irreducible E-modules. Going back to A we see that the identity $1 \in A$ can be decomposed $1 = e_1 + \cdots + e_n$ so that the $B_i = Ae_i$ are irreducible A-submodules (i.e., left ideals) and $A = B_1 + \cdots + B_n$ is a direct sum.

(2) Let $B_i = A e_i$ be an irreducible left ideal of A as above and let V be an A-module. If $X \in V$ is such that $e_i X \neq 0$, then $B_i X \neq \{0\}$ and $B_i X$ is an A-module. The map

$$\phi : B_i \to B_i X : b \to bX$$

is an A-module homomorphism and $\ker(\phi)$ is an A-submodule of B_i. Since B_i is an irreducible A-module and $B_i X \neq \{0\}$ we have ϕ is an isomorphism; that is, $B_i X$ is an irreducible A-submodule of V.

(3) Let A be an associative algebra with identity 1 and let V be an A-module. Then V is called a **unital** A-module if $1X = X$ for all $X \in V$.

Let $A = B_1 \oplus \cdots \oplus B_n$ as in remark (1), let V be a unital A-module, and let $0 \neq X \in V$. Then $X = 1X = e_1 X + \cdots + e_n X$ so there exists e_i with $e_i X \neq 0$. Thus $B_i X$ is an irreducible A-module and $X \in B_1 X + \cdots + B_n X$, a sum of irreducible A-modules. This implies V is a completely reducible A-module. Thus we leave as an exercise: Let $\{W_j : j \in \alpha\}$ be a family of irreducible A-submodules of V so that every $X \in V$ can be expressed as a finite sum of elements from the W_j's. Then V is a direct sum of irreducible A-submodules.

For a semisimple algebra A we combine the above remarks to obtain the following.

Proposition 9.16 Let A be a finite-dimensional semisimple associative algebra over K and let V be a finite-dimensional A-module. Then V is a completely reducible A-module.

PROOF Since $1 \in A$ let $V_0 = \{Z \in V : 1Z = 0\}$. Then V_0 is an A-submodule of V such that $AV_0 = \{0\}$. Next note any $X \in V$ can be written

$$X = 1X + (X - 1X)$$

where $1(1X) = 1X$ and $1(X - 1X) = 0$; that is, $X = Y + Z$ where $Z \in V_0$ and $Y \in V_1 = \{U \in V : 1U = U\}$. This gives the direct sum $V = V_1 + V_0$ where V_1 is a unital A-module and therefore completely reducible. Now choosing a basis for V_0 we can write $V_0 = KX_1 + \cdots + KX_r$ as a direct sum of one-dimensional A-modules which are irreducible. Combining these decompositions, V is completely reducible.

REMARK (4) Concerning complexification, we shall use bilinear forms in Chapter 12 to easily show a (finite-dimensional) associative algebra A over R is semisimple if and only if the associative algebra $C \otimes_R A$ over C is semisimple. This with the preceding results shows that if $1 \in A$ and V is an A-module, then V is a completely reducible A-module if and only if $C \otimes_R V$ is a completely reducible $C \otimes_R A$-module.

6. Composition Algebras

We will now construct some nonassociative algebras which will be very useful in describing certain simple Lie algebras in Chapters 13 and 14. This material also yields some interesting applications of the previous material in this chapter.

Definition 9.17 **A composition algebra** $\mathscr{C}$ over a field K of characteristic 0, is a nonassociative algebra $\mathscr{C}$ over K with an identity element denoted by 1 and a map $N : \mathscr{C} \to K$ called the **norm** of $\mathscr{C}$ such that the following three properties hold:

(a) $N(aX) = a^2 N(X)$ for any $a \in K$ and any $X \in \mathscr{C}$;
(b) $N(XY) = N(X)N(Y)$ for any $X, Y \in \mathscr{C}$;
(c) $B(X, Y) = [N(X + Y) - N(X) - N(Y)]/2$ for all $X, Y \in \mathscr{C}$ defines a nondegenerate symmetric bilinear form on $\mathscr{C}$.

Exercises (1) Show that $N(1) = 1$ and $B(X, X) = N(X)$ for all $X \in \mathscr{C}$ where $\mathscr{C}$ is any composition algebra.

(2) Suppose $\lambda \in K$ is not a square in K. Let $\mathscr{C} = K(\lambda^{1/2})$ be the quadratic field extension of K and define $N(a + b\lambda^{1/2}) = a^2 - b^2\lambda$. Show that $\mathscr{C}$ is a composition algebra.

(3) Show that the associative algebra of 2×2 matrices over K is a composition algebra if we define $N(X) = \det(X)$.

Definition 9.18 For any composition algebra $\mathscr{C}$ and any $X \in \mathscr{C}$ define $\bar{X} = 2B(X, 1)1 - X$. Then $\bar{X}$ is called the **conjugate** of X and the overbar is called the **involution** of $\mathscr{C}$. Notice that $B(X, 1)1 = (X + \bar{X})/2$ and that if we write $X = a1 + Y$ with $B(Y, 1) = 0$, then $\bar{X} = a1 - Y$.

Proposition 9.19 For any X, Y, Z, W in a composition algebra $\mathscr{C}$ we have

(a) $B(XY, XZ) = N(X)B(Y, Z) = B(YX, ZX)$;
(b) $B(XY, WZ) + B(XZ, WY) = 2B(X, W)B(Y, Z)$;
(c) $X^2 - 2B(X, 1)X + N(X)1 = 0$;
(d) $B(XY, Z) = B(Y, \bar{X}Z) = B(X, Z\bar{Y})$.

PROOF (a) From Definition 9.17 we see

$$\begin{aligned} 2B(XY, XZ) &= N(XY + XZ) - N(XY) - N(XZ) \\ &= N(X(Y + Z)) - N(XY) - N(XZ) \\ &= N(X)(N(Y + Z) - N(Y) - N(Z)) = N(X)2B(Y, Z). \end{aligned}$$

This kind of proof is routine and is referred to as "linearizing" equation 9.17(b) with respect to Y.

(b) This is proved by linearizing (a) with respect to X.

(c) We note that

$$\begin{aligned} B(X^2 - 2B(X, 1)X + N(X)1, Z) &= B(X^2, Z) - 2B(X, 1)B(X, Z) \\ &\quad + N(X)B(1, Z) \\ &= B(X^2, Z \cdot 1) + B(X \cdot 1, Z \cdot X) \\ &\quad - 2B(X, Z)B(X, 1) = 0. \end{aligned}$$

Both (a) and (b) above were used in the computation. Now the formula in (c) follows from the nondegeneracy of $B(X, Y)$.

(d) From (b) we have

$$B(XY, Z) + B(XZ, Y) = 2B(X, 1)B(Y, Z),$$

and thus

$$B(XY, Z) = B(Y, 2B(X, 1)Z - XZ) = B(Y, \bar{X}Z),$$

and finally the other parts of (a) and (d) which were not proved follow from obvious symmetries.

Exercise (4) Prove that the following formulas hold for any X and Y in a composition algebra:

(i) $X\overline{X} = \overline{X}X = N(X)1$;
(ii) $\overline{XY} = \overline{Y}\overline{X}$
(iii) $\overline{\overline{X}} = X$

Properties (ii) and (iii) are often used to define an involution.

Definition 9.20 Given a composition algebra $\mathscr{C}$ over K and any $0 \neq a \in K$, define a new algebra $\mathscr{C}\langle a\rangle$ with underlying vector space $\mathscr{C} \times \mathscr{C}$ and a multiplication and quadratic form N' given by the following **Cayley–Dickson formulas** for $X, Y, Z, W \in \mathscr{C}$

$$(X, Y)(Z, W) = (XZ + a\overline{W}Y, WX + Y\overline{Z}),$$
$$N'((X, Y)) = N(X) - aN(Y).$$

We will shortly state a theorem which shows that deciding when the algebra $\mathscr{C}\langle a\rangle$ is a composition algebra is a way of classifying all composition algebras.

REMARKS (1) Rather than using pairs (X, Y), one can describe $\mathscr{C}\langle a\rangle$ by using a symbol, say u. Then $\mathscr{C}\langle a\rangle = \mathscr{C} + \mathscr{C}u = \{X + Yu : X, Y \in \mathscr{C}\}$ with $N'(X + Yu) = N(X) - aN(Y)$ and $X(Wu) = (WX)u$, $(Yu)Z = (Y\overline{Z})u$ and $(Yu)(Wu) = a\overline{W}Y$.

(2) Since K is a composition algebra if we define $N(a) = a^2$, we can consider $K\langle\alpha\rangle$ for any $0 \neq \alpha \in K$ and if this is a composition algebra, consider $K\langle\alpha, \beta\rangle = K\langle\alpha\rangle\langle\beta\rangle$ and continue until one obtains an algebra which is not a composition algebra. The $K\langle\alpha\rangle$'s are called **quadratic algebras**, the $K\langle\alpha, \beta\rangle$'s are called **(generalized) quaternion algebras** and the $K\langle\alpha, \beta, \gamma\rangle$'s are called **(generalized) Cayley algebras**. We are justified in giving names to these algebras by Hurwitz's theorem, the proof of which can be found in the work of Schafer [1966, Chap. 3].

Theorem 9.21 (Hurwitz) All quadratic, quaternion, and Cayley algebras are composition algebras. If $\mathscr{C}$ is a composition algebra over K of dimension greater than 1, then $\mathscr{C}$ must be isomorphic to some quadratic, quaternion, or Cayley algebra over K. Quadratic algebras are commutative and associative, quaternion algebras are associative but not commutative, and Cayley algebras are neither commutative nor associative.

REMARK (3) Proposition 9.19(c) can be used to show that any algebras isomorphism of two composition algebras also leaves the norms and bilinear forms invariant. Thus the isomorphisms referred to in Theorem 9.21 can be thought of as isomorphisms of only the algebra structures or as norm-invariant isomorphisms.

Example (1) We often refer to $R\langle -1, -1\rangle$ as "the" quaternion numbers. It can be described as the real associative algebra with basis 1, u, v, uv, where $u^2 = v^2 = -1$, $vu = -uv$, and $N(a + bu + cv + duv) = a^2 + b^2 + c^2 + d^2$. Then $R\langle -1, -1\rangle$ is a division algebra.

Definition 9.22 A composition algebra $\mathscr{C}$ over K is said to be a **split composition algebra** if it possesses zero divisors or equivalently if it is not a division algebra.

Proposition 9.23 The following statements about a composition algebra $\mathscr{C}$ over K are equivalent:

(a) $\mathscr{C}$ is a split composition algebra;
(b) there exists a nonzero $X \in \mathscr{C}$ with $N(X) = 0$;
(c) there exists an $X \in \mathscr{C}$ with $B(X, 1) = 0$ and $N(X) = -1$.

PROOF (a) implies (b) The algebra $\mathscr{C}$ split implies there exist nonzero $X, Y \in \mathscr{C}$ with $XY = 0$ and $0 = N(XY) = N(X)N(Y)$ so $N(X) = 0$ or $N(Y) = 0$.

(b) implies (c) Suppose $N(X) = 0$ for $0 \neq X \in \mathscr{C}$. If $B(1, X) \neq 0$, set $Y = 1 - B(1, X)^{-1}X$ and compute that $B(Y, 1) = 0$ and $N(Y) = -1$. If $B(1, X) = 0$, then from the nondegeneracy of the bilinear form on $\mathscr{C}$ we must have that $\mathscr{C}$ is four- or eight-dimensional and there exists some $Z \in \mathscr{C}$ with $B(1, Z) = 0$ and $B(X, Z) = 2$. Let $Y = Z - (N(Z) + 1)X$ and compute $B(Y, 1) = 0$ and $N(Y) = -1$.

(c) implies (a) If $X \in \mathscr{C}$ with $B(X, 1) = 0$ and $N(X) = -1$, then, by Proposition 9.19(c), $X^2 = 1$; so $(X + 1)(X - 1) = 0$ and so $X + 1$ is a zero divisor and $\mathscr{C}$ is a split composition algebra.

Proposition 9.24 Over any field K there exist split quadratic, quaternion, and Cayley algebras. Any two split composition algebras over K of the same dimension are isomorphic.

PROOF We see $K\langle 1\rangle$, $K\langle 1, 1\rangle$, and $K\langle 1, 1, 1\rangle$ are split algebras from Proposition 9.23. Thus it suffices to show that any split composition algebra over K must be isomorphic to one of these three algebras.

Assume $\mathscr{C}$ is split. Proposition 9.23(c) shows that there exists an $U \in \mathscr{C}$ with $B(1, U) = 0$ and $N(U) = -1$. Since $U^2 = 1$, $\mathscr{A} = K1 + KU$ is a subalgebra of $\mathscr{C}$ and it is trivial to check that $a1 + bU \to (a, b)$ is an isomorphism from $\mathscr{A}$ onto $K\langle 1 \rangle$.

If $\mathscr{C}$ has dimension greater than 2, we must continue. Choose any $0 \neq Y \in \mathscr{C}$ with $B(X, Y) = 0$ for all $X \in \mathscr{A}$. If $X \in \mathscr{A}$ with $N(X) = 0$, then $N(XY) = 0$ and by Proposition 9.19(d) $B(Z, XY) = 0$ for all $Z \in \mathscr{A}$. Now repeating an argument used in the second part of the proof of 9.23 there exist a $V \in \mathscr{C}$ with $B(Z, V) = 0$ for all $Z \in \mathscr{A}$ and $N(V) = -1$. It is claimed that $\mathscr{A} + \mathscr{A}V$ is isomorphic to $\mathscr{A}\langle 1 \rangle \cong K\langle 1, 1 \rangle$. Using Proposition 9.19(d) it is easy to show that $N(X + ZV) = N(X) - N(Z)$ as required so we need only check the nontrivial multiplication formulas.

Assume $X, Y, Z \in \mathscr{A}$ in this paragraph. Since $B(X, YV) = B(X\bar{Y}, V) = 0$ and $B(XV, YV) = -B(X, Y)$ it is clear that the bilinear form restricted to $\mathscr{A}V$ is nondegenerate and $\mathscr{A}V$ is orthogonal to $\mathscr{A}$. For any $W \in \mathscr{C}$, $X, Y \in \mathscr{A}$ and V as above

$$B(YV, WX) + B(YX, WV) = 2B(Y, W)B(V, X) = 0$$

so $B((YV)\bar{X}, W) = B((YX)V, W)$ for all $W \in \mathscr{C}$ and $(YX)V = (YV)\bar{X}$. Applying the involution to both sides of this equation yields $X(YV) = (YX)V$ which is one of the three multiplication formulas required.

The second formula $(XV)Y = (X\bar{Y})V$ follows easily from the first and involution formulas.

Finally, for any $W \in \mathscr{C}$

$$B(W(YV), V\bar{X}) + B(W\bar{X}, V(YV)) = 2B(W, V)B(YV, \bar{X}) = 0$$

so $(XV)(YV) = (V(YV))X$. However $B(W, V(YV)) = -B(VW, V\bar{Y}) = B(W, \bar{Y})$ so $V(YV) = \bar{Y}$ and $(XV)(YV) = \bar{Y}X$.

This completes the proof that $\mathscr{B} = \mathscr{A} + \mathscr{A}V$ is a subalgebra of $\mathscr{C}$ isomorphic $\mathscr{A}\langle 1 \rangle$ which is isomorphic to $K\langle 1, 1 \rangle$. If $\mathscr{C}$ is eight-dimensional, we now repeat the above argument word for word to complete the proof of the proposition.

Exercises (5) Show that the algebra of 2×2 matrices over K is a split quaternion algebra. Also show that $K\langle \alpha \rangle \cong K(\alpha^{1/2})$ if $0 \neq \alpha \in K$ is not a square in K and $K\langle \alpha \rangle$ is split otherwise.

(6) Show that if $\mathscr{C}$ is a composition algebra over K and L is a field extension of K, then $\mathscr{C}(L) = L \otimes_K \mathscr{C}$ is a composition algebra over L where $N(a \otimes X) = a^2 N(X)$ for $a \in L$ and $X \in \mathscr{C}$. Also show that if $\mathscr{C}$ is split, then so is $\mathscr{C}(L)$.

(7) Suppose that $\mathscr{C}$ is a division (nonsplit) composition algebra of dimension greater than 1, over K and that $N(X) = \alpha \neq 0$ for some $X \in \mathscr{C}$

with $B(X, 1) = 0$. Show that $\mathscr{C}(K((-\alpha)^{1/2})) = K((-\alpha)^{1/2}) \otimes_K \mathscr{C}$ is a split composition algebra over $K((-\alpha)^{1/2})$. Use this result to show that if $\mathscr{C}_1$ and $\mathscr{C}_2$ are any two composition algebras of the same dimension over K, there exists a field extension L of degree at most 4 over K so that $\mathscr{C}_1(L) \cong \mathscr{C}_2(L)$.

Example (2) The following matrix-type description of a split Cayley algebra is sometimes useful. Let K be any field of characteristic 0 and let $\mathscr{V}$ denote the set of three-dimensional column vectors over K. For any $u, v \in \mathscr{V}$ let $u \times v$ and (u, v) denote the usual cross product and inner product, respectively. Let

$$\mathscr{C} = \left\{ \begin{bmatrix} a & u \\ v & b \end{bmatrix} : a, b \in K, u, v \in \mathscr{V} \right\}$$

and define

$$\begin{bmatrix} a & u \\ v & b \end{bmatrix} \begin{bmatrix} c & w \\ z & d \end{bmatrix} = \begin{bmatrix} ac - (u, z) & aw + du + v \times z \\ cv + bz + u \times w & bd - (v, w) \end{bmatrix}$$

and

$$N\left(\begin{bmatrix} a & u \\ v & b \end{bmatrix}\right) = ab + (u, v)$$

for all $a, b, c, d \in K$ and $u, v, w, z \in \mathscr{V}$.

To show that $\mathscr{C}$ is a composition algebra we must verify the three formulas in Definition 9.17 for composition algebras. Formulas (a) and (c) are trivial to verify and (b) is equivalent to

$$\begin{aligned}(ab + (u, v))(cd + (w, z)) = {} & (ac - (u, z))(bd - (v, w)) \\ & + (aw + du + v \times z, cv + bz + u \times w)\end{aligned}$$

It is easy to check that this formula follows from two well-known vector equations, namely $(u \times v, u) = (u \times v, v) = 0$ for all $u, v \in \mathscr{V}$ and $(u \times w, v \times z) = (u, v)(w, z) - (u, z)(v, w)$ for all $u, v, w, z \in \mathscr{V}$. Finally it is very easy to choose nonzero elements of $\mathscr{C}$ of norm 0 so $\mathscr{C}$ is a split eight-dimensional composition algebra; that is, a split Cayley algebra.

Exercise (8) Let $\mathscr{C}$ be any composition algebra over K, let L be a field extension of K of finite degree, and let $\mathscr{D}(\mathscr{C})$ denote the derivation algebra of $\mathscr{C}$. Show that $\mathscr{D}(\mathscr{C}(L)) \cong (\mathscr{D}(\mathscr{C}))(L)$.

REMARK (4) One of the important properties of quaternion and Cayley algebras is that they are simple nonassociative algebras. This fact can be used to show that all derivations of these two types of algebras are inner; that is, for these algebras $\mathscr{D}(\mathscr{C})$ is contained in $L(\mathscr{C})$ which is the subalgebra of $gl(\mathscr{C})$

generated by the linear transformations $L(X) : Y \to XY$ and $R(X) : Y \to YX$. The proof of these results can be found in the work of Schafer [1966, Chap. 3] or Section 12.5, exercise (3).

Exercise (9) Let $\mathscr{Q}$ be a quaternion algebra over K and for $X \in \mathscr{Q}$ define $D(X) = L(X) - R(X)$. Assuming the results mentioned in the above remark show that $\mathscr{D}(\mathscr{Q}) = \{D(X) : X \in \mathscr{Q}, B(X, 1) = 0\}$ and that $[D(X), D(Y)] = D(XY - YX)$. Also show that for any $\mathscr{Q}$ there exists a field extension L of K of at most degree 2 over K such that $(\mathscr{D}(\mathscr{Q}))(L) \cong sl(2, L)$.

Example (3) It is much more difficult to describe the derivation algebra of a Cayley algebra but it is not quite so difficult for the split Cayley algebra as described in the previous example. For any $A \in sl(3, K)$ and any $x, y \in \mathscr{V}$ define $D(A, x, y) : \mathscr{C} \to \mathscr{C}$ by

$$D(A, x, y) : \begin{bmatrix} a & u \\ v & b \end{bmatrix} \to \begin{bmatrix} (x, v) + (y, u) & Au + (a - b)x - y \times v \\ -A^t v + (a - b)y + x \times u & -(x, v) - (y, u) \end{bmatrix}.$$

A long computation involving a few vector formulas verifies that $D(A, x, y) \in \mathscr{D}(\mathscr{C})$ and a similarly easy but unpleasantly long computation shows that $[D(A_1, x_1, y_1), D(A_2, x_2, y_2)] = D(A_3, x_3, y_3)$, where

$$\begin{aligned} A_3 &= [A_1, A_2] + 3x_1 y_2{}^t - 3x_2 y_1{}^t + ((x_2, y_1) - (x_1, y_2))I, \\ x_3 &= A_1 x_2 - A_2 x_1 - 2y_1 \times y_2, \\ y_3 &= -A_1{}^t y_2 + A_2{}^t y_1 + 2x_1 \times x_2. \end{aligned}$$

Some of these computations can be found in Schafer's book [1966] which also has a proof that $\mathscr{D}(\mathscr{C}) = \{D(A, x, y) \mid A \in sl(3, K), x, y \in \mathscr{V}\}$.

Using exercise (8) we can now conclude that for any composition algebra $\mathscr{C}$, $\mathscr{D}(\mathscr{C})$ must be fourteen dimensional because we can extend the base field to say L so that $\mathscr{C}(L)$ is split and then notice that we can compute the dimension of $\mathscr{D}(\mathscr{C}(L))$ from the description above. Also notice that for any $X \in \mathscr{C}$ and $D \in \mathscr{D}(\mathscr{C})$ we have $B(D(X), 1) = 0$ since this is easy to verify in the split case. Since $D(1) = 0$ we can conclude that $\mathscr{D}(\mathscr{C})$ also can be thought of as the set of derivations of $\mathscr{C}$ acting on the elements of trace 0 that is acting on $\mathscr{C}_0 = \{X \in \mathscr{C} : B(X, 1) = 0\}$. This may be restated as: $\mathscr{C}_0$ is an invariant submodule of $\mathscr{C}$ or the action of $\mathscr{D}(\mathscr{C})$ on $\mathscr{C}_0$ gives a seven-dimensional representation of $\mathscr{D}(\mathscr{C})$. We will see in Chapters 13 and 14 that $\mathscr{D}(\mathscr{C})$ is simple.

Exercise (10) If $\mathscr{C}$ is any Cayley algebra, show that $\mathscr{D}(\mathscr{C})$ acts irreducibly on $\mathscr{C}_0$. Notice that you may assume the $\mathscr{C}$ is split.

CHAPTER 10

SOLVABLE LIE GROUPS AND ALGEBRAS

We now start the structural development of Lie groups and algebras. First we define a Lie group to be solvable if it is solvable as an abstract group. Then using the "derivative" of these results we discuss solvable Lie algebras. Thus we show that a connected Lie group is solvable if and only if its Lie algebra is solvable. Finally we discuss Lie's theorem which involves finding a common characteristic vector for a solvable Lie algebra of endomorphisms acting on a complex vector space. This eventually yields that the matrices representing a solvable Lie algebra of endomorphisms acting on a complex vector space can be put into triangular form by using a suitable basis of the vector space. Once again, all fields in this chapter will be assumed to be of characteristic zero.

1. Solvable Lie Groups

Let G be an abstract group and let A and B be subgroups of G. Then we have the following notation.

(1) We denote by (A, B) the subgroup of G generated by all elements $xyx^{-1}y^{-1}$ for $x \in A$, $y \in B$.

(2) If A is a normal subgroup of G we write $A \lhd G$ or $G \rhd A$.

Note that if A and B are normal subgroups of G, then (A, B) is a normal subgroup of G.

Definition 10.1 Let $G^{(1)} = (G, G)$ and define by induction $G^{(k+1)} = (G^{(k)}, G^{(k)})$. Then we have the sequence of normal subgroups

$$G \rhd G^{(1)} \rhd G^{(2)} \rhd \cdots.$$

Thus G is **solvable** if this sequence is finite and terminates at $\{e\}$; that is, there exists n so that $G^{(n)} = \{e\}$ and G is called solvable of **length** n.

From results of Lang [1965] we have the following theorem:

Theorem 10.2 Let G be an abstract group. Then the following are equivalent.

(a) The group G is solvable.

(b) There is a finite sequence of subgroups $G = G_0 \rhd G_1 \rhd G_2 \cdots \rhd G_n = \{e\}$ such that G_k/G_{k+1} is commutative for $k = 0, 1, \ldots, n-1$.

Proof First we observe that by induction each $G^{(k)}$ is a normal subgroup of G. Now assume (a). Then note from the definition of (G, G) that $G/G^{(1)}$ is commutative and by induction and definition, $G^{(i)}/G^{(i+1)}$ is also commutative. Thus we have (b) by taking $G_k = G^{(k)}$. Conversely, assume we have a descending sequence

$$G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{e\}$$

with G_i/G_{i+1} commutative. Then G/G_1 being commutative implies $xyx^{-1}y^{-1}G_1 = eG_1$ which yields $G_1 \supset G^{(1)}$. Now assume

$$G_k \supset G^{(k)}$$

Then since G_k/G_{k+1} is commutative we see

$$G_{k+1} \supset (G_k)^{(1)} \supset (G^{(k)}, G^{(k)}) = G^{(k+1)}.$$

However, since $G_n = \{e\}$ we see $G^{(n)} = \{e\}$ which gives (a).

Corollary 10.3 (a) A subgroup H of a solvable group G is solvable.

(b) If G is a solvable group of length n and H a normal subgroup, then G/H is a solvable group of length less than or equal to n.

(c) If G is a group and H is a normal solvable subgroup of length n such that G/H is solvable of length m, then G is solvable of length less than or equal to $n + m$.

Proof (a) We just note that by induction $H^{(k)} \subset G^{(k)}$.

(b) Let $\bar{G} = G/H$. Then by induction we see that $(\bar{G})^{(k)} = \overline{G^{(k)}}$ using $\pi : G \to G/H = x \to \bar{x} = xH$ is a homomorphism. Thus the series for G yields the series $\bar{G} \rhd \bar{G}^{(1)} \rhd \cdots \rhd \bar{G}^{(n)} = \{eH\}$.

(c) Note that from the series $\bar{G} \rhd \bar{G}^{(1)} \rhd \cdots \rhd \bar{G}^{(n)} = \{eH\}$ we obtain $G \rhd G^{(1)} \rhd \cdots \rhd G^{(n)}$ and $G^{(n)} \subset H$. However, since H is solvable we have $H \rhd H^{(1)} \rhd H^{(2)} \rhd \cdots \rhd \{e\}$ and we put these two series together to see that G is solvable.

Definition 10.4 Let G be a Lie group. Then G is a **solvable Lie group** if G is solvable as an abstract group.

Theorem 10.5 Let G be a Lie group. Then the following are equivalent.

(a) The Lie group G is solvable.
(b) There exists a finite sequence of subgroups,

$$G = G_0 \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_n = \{e\}$$

such that each G_k is a closed Lie subgroup of G with G_k/G_{k+1} commutative for $k = 0, 1, \ldots, n-1$.

(c) There exists a finite sequence of closed Lie subgroups

$$G = G_0 \rhd G_1 \rhd \cdots \rhd G_r = \{e\}$$

such that for $k = 0, \ldots, r-1$, we have G_k/G_{k+1} is a connected one-dimensional group or a discrete group.

PROOF To show (a) implies (b), we recall that if H is a normal subgroup of G, then its closure $\bar{H}$ is a closed normal subgroup of G. Next assume G is solvable so we obtain the series

$$G \rhd G^{(1)} \rhd G^{(2)} \rhd \cdots \rhd G^{(m)} = \{e\},$$

and let

$$G_0 = G \qquad \text{and} \qquad G_k = \overline{G^{(k)}}.$$

Then the G_k are closed normal subgroups (and, therefore, Lie subgroups) such that $G_0 \rhd G_1 \rhd \cdots \rhd G_m = \{e\}$. Furthermore G_k/G_{k+1} is a Lie group which is commutative, for let $\pi : G \to G/G_{k+1}$. Then since $\pi(G^{(k)})$ is commutative, we have

$$\overline{\pi(G^{(k)})}$$

is commutative. However, since π is continuous,

$$\pi(G_k) = \pi(\overline{G^{(k)}}) \subset \overline{\pi(G^{(k)})}$$

so that $G_k/G_{k+1} = \pi(G_k)$ is commutative.

The converse (b) implies (a) is clear. Also (c) implies (b) is clear, so it remains to show (b) implies (c). Thus let H_k be the connected component of the commutative Lie group G_k/G_{k+1}. Then by the results outlined in the exercise (1), Section 6.5, we have for the G_k in (b)

$$G_k/G_{k+1} = H_k \times D_k$$

where D_k is a discrete group, and $H_k \cong R^{q(k)} \times T^{p(k)}$ by Theorem 6.20.

Next let $\pi : G_k \to G_k/G_{k+1}$. Then in the series for (b) we replace each of the terms G_k by the series [with $p = p(k)$, $q = q(k)$]

$$\pi^{-1}(R^p \times T^q \times D_k) \rhd \pi^{-1}(R^{p-1} \times T^q \times D_k) \rhd \cdots \rhd \pi^{-1}(T^q \times D_k)$$
$$\rhd \pi^{-1}(T^{q-1} \times D_k) \rhd \cdots \rhd \pi^{-1}(D_k).$$

Thus we obtain the series in (c).

Exercise (1) Let V be a real vector space of dimension m and regard $GL(V)$ as the set of all nonsingular $m \times m$ real matrices. Let H be the subset of all matrices of $GL(V)$ of the form

$$\begin{bmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{mm} \end{bmatrix}$$

where $a_{jj} \neq 0$ and $*$ arbitrary real numbers. Show H is a solvable Lie subgroup of $GL(V)$.

2. Solvable Lie Algebras and Radicals

Let g be a finite-dimensional Lie algebra over a field K and let h, k be subspaces of g. Then we shall use the following notation.

(1) We denote by $[hk]$ the subspace of g generated by all products $[xy]$ for $x \in h$ and $y \in k$. In particular, $g^{(1)} = [gg]$ is a subalgebra of g.

(2) If h is an ideal of g, then we write $g \rhd h$ or $h \lhd g$. In particular, note $g \rhd g^{(1)}$. A Lie algebra g is **abelian** or **commutative** if $g^{(1)} = \{0\}$.

Definition 10.6 Let g be a finite-dimensional Lie algebra over K, set $g^{(1)} = [gg]$, and define by induction

$$g^{(k+1)} = [g^{(k)}g^{(k)}].$$

From the Jacobi identity for g we obtain

$$g \rhd g^{(1)} \rhd g^{(2)} \rhd \cdots$$

and we call g **solvable** if there exists n with $g^{(n)} = \{0\}$. The smallest such n is called the **length** of the solvable algebra g.

Theorem 10.7 Let g be a finite-dimensional Lie algebra over K. Then the following are equivalent.

(a) The algebra g is solvable.

(b) There exists a sequence of subalgebras $g = g_0 \rhd g_1 \rhd \cdots \rhd g_r = \{0\}$ so that the quotient algebra g_k/g_{k+1} is commutative. Each g_k can be taken to be an ideal in g.

(c) There exists a finite sequence of subalgebras $g = g_0 \rhd g_1 \rhd \cdots \rhd g_s = \{0\}$ such that dim g_k/g_{k+1} is 1. In general g_k is not an ideal in g but only in g_{k-1}.

PROOF The equivalence of (a) and (b) is similar to those for groups in Section 10.1. Thus, for example, if g is solvable, then take $g_k = g^{(k)}$ for $k = 1, \ldots, r$ to obtain the sequence in (b) and also note $[g^{(k)}g^{(k)}] = g^{(k+1)}$ so the desired quotient algebra is commutative.

Next assume (c) where we have $g_k/g_{k+1} = K\overline{X} = KX + g_{k+1}$ since g_k/g_{k+1} is one dimensional. Then since $[\overline{X}\overline{X}] = 0$ we have g_k/g_{k+1} is a commutative Lie algebra. Thus (c) implies (b). Conversely, if the sequence in (b) is such that $g_k/g_{k+1} = KX_1 + \cdots + KX_r + g_{k+1}$ is commutative, then each subspace $\bar{h}(i) = KX_1 + \cdots + KX_i + g_{k+1}$ is an ideal in g_k/g_{k+1} for $i = 1, \ldots, r$. Thus the corresponding subspace $h(i)$ generated by $\{X_1, \ldots, X_i\} \cup g_{k+1}$, where $X_i + g_{k+1} = \overline{X}_i \in g_k/g_{k+1}$, is an ideal in g_k. Thus we obtain a sequence

$$g_k = h_r \rhd h_{r-1} \rhd \cdots \rhd h_1 \rhd g_{k+1}$$

so that the quotient ideals are one dimensional and this yields (c).

The proof of the following is similar to Corollary 10.3.

Corollary 10.8 Let g be a Lie algebra containing the Lie subalgebra h.

(a) If g is solvable, then h is solvable.

(b) If g is solvable and h an ideal of g, then g/h is solvable of length less than or equal to the length of g.

(c) If h is a solvable ideal of g such that g/h is solvable, then g is solvable.

Exercises (1) Let g denote the set of $m \times m$ matrices of the form

$$\begin{bmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{mm} \end{bmatrix}$$

where a_{ij} are arbitrary in K. Show g is a solvable Lie subalgebra of $gl(V)$.

(2) Prove the following isomorphism theorems for Lie algebras.

(i) Let $f: g \to \bar{g}$ be a homomorphism of the Lie algebra g onto the Lie algebra $\bar{g}$ and let $k = \ker(f)$. Then $\bar{h}$ is an ideal of $\bar{g}$ if and only if the inverse image $f^{-1}(\bar{h}) \equiv h$ is an ideal of g such that $h \supset k$. If this is the case, we have $g/h \cong \bar{g}/\bar{h}$; that is, $g/h \cong (g/k)/(h/k)$.

(ii) Let h and k be ideals of the Lie algebra g and let $f: g \to g/k$ be the natural homomorphism. Then $h + k = f^{-1}(f(h))$ and $(h + k)/k \cong h/h \cap k$.

(iii) Also show that the sum $h + k$ and product $[hk]$ of the ideals h and k are again ideals of g.

(3) Let g be a solvable Lie algebra of dimension n over K. Show that g is a semidirect sum of an ideal h of dimension $n - 1$ and a one-dimensional subalgebra.

(4) Show that if g is a solvable Lie algebra over K, then $P \otimes_K g$ is a solvable Lie algebra over the algebraic closure P of K.

Theorem 10.9 Let G be a Lie group with Lie algebra g.

(a) If G is solvable, then g is solvable.
(b) If G is connected and g is solvable, then G is solvable.

PROOF (a) If G is a solvable Lie group, then we have a sequence $G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{e\}$ with each G_k a closed normal Lie subgroup so that G_k/G_{k+1} is commutative. Then we obtain the corresponding sequence $g = g_0 \rhd g_1 \rhd \cdots \rhd g_n = \{0\}$ of ideals of g so that g_k/g_{k+1} is a commutative Lie algebra.

(b) If G is connected and g is solvable, then we shall show G is solvable by induction on the length of g. Thus let $g \rhd g^{(1)} \rhd \cdots \rhd g^{(n-1)} \rhd g^{(n)} = \{0\}$ be the sequence for g and let K be the Lie subgroup of G generated by $\exp g^{(n-1)}$. Then K is a commutative normal subgroup of G (since $g^{(n-1)}$ is a commutative ideal of g) and its closure $\bar{K} = H$ is also a commutative normal Lie subgroup of G. Now let h be the Lie algebra of H. Then h is a commutative ideal of g and $g^{(n-1)} \subset h$. From this we have [exercise (2) above], $g/h \cong (g/g^{(n-1)})/(h/g^{(n-1)})$ and since $g/g^{(n-1)}$ is solvable of length less than or equal to $n - 1$ we have g/h is solvable of length less than or equal to $n - 1$ (Corollary 10.8).Thus by the induction hypotheses G/H is solvable and since H is solvable, we have G is solvable using Corollary 10.3.

Exercise (5) If G is a connected solvable Lie group, then there exists a sequence $G = G_0 \rhd G_1 \rhd \cdots \rhd G_m = \{e\}$ where all the G_k are closed connected Lie subgroups such that G_k/G_{k+1} are one dimensional. In particular, G contains a connected solvable normal subgroup H with $\dim H = \dim G - 1$.

Lemma 10.10 Let g be a finite-dimensional Lie algebra over K. Then there exists a unique maximal solvable ideal of g: namely the sum of the solvable ideals of g. This maximal solvable ideal is called the **radical** of g and is denoted by r. Moreover g/r is $\{\bar{0}\}$ or contains no proper solvable ideals; that is, the radical of g/r is $\{\bar{0}\}$.

PROOF Let h and k be solvable ideals of g. Then the vector subspace $h + k$ is an ideal of g. Now by the above exercise (2) we see $(h + k)/k \cong h/(h \cap k)$ and since $h \cap k \subset h$ is solvable we have $h/(h \cup k)$ is solvable. Thus we have $(h + k)/k$ is solvable and k is solvable so that by Corollary 10.8, $h + k$ is solvable. Thus since g is finite dimensional, the solvable ideal of maximum dimension is unique and by the above, contains every solvable ideal of g; denote this maximal solvable ideal by r.

Next let $\bar{h} = h/r$ be a solvable ideal of $\bar{g} = g/r$ where h is some ideal of g with $h \supset r$. Then since h/r is solvable and r is solvable we have by Corollary 10.8 that h is solvable. Thus $h \subset r$, so that $\bar{h} = \{\bar{0}\}$.

Definition 10.11 Let G be a Lie group with Lie algebra g and let r be the radical of g. Then we define the **radical** of G, $R = \operatorname{rad} G$, to be the connected Lie subgroup of G whose Lie algebra is $r = \operatorname{rad} g$.

Proposition 10.12 Let G be a Lie group with radical R. Then R is closed and R is the maximal solvable normal connected Lie subgroup of G.

PROOF Let $\bar{R}$ denote the closure of R. Then $\bar{R}$ is a normal, solvable Lie subgroup (since it is closed). Thus its Lie algebra $\bar{r}$ is solvable (Theorem 10.9) so that $r = \bar{r}$ and consequently $\bar{R} = R$; that is, R is a closed, normal, solvable Lie subgroup of G. The fact that R is maximal among connected Lie subgroups with these properties also uses the maximality of r.

Corollary 10.13 The radical of G/R equals $\{eR\}$.

Exercise (6) Show that the radical of a Lie algebra g is the smallest ideal h of g such that the radical of g/h is $\{\bar{0}\}$; that is, if h satisfies this condition, then $r \subset h$.

Definition 10.14 (a) A finite-dimensional Lie algebra is called **semisimple** if it has no proper solvable ideals. Thus g is semisimple if and only if $r = \{0\}$. Similarly a Lie group G is **semisimple** if its radical $R = \{e\}$.

(b) A Lie group G is **simple** if its Lie algebra g is simple. That is, $[gg] \neq \{0\}$ and g has no proper ideals.

We shall eventually show that a semisimple Lie algebra over a field of characteristic 0 is a direct sum of simple Lie algebras which are ideals. Consequently many problems involving semisimple Lie groups can be done in terms of simple Lie algebras.

Exercise (7) Show that the center of a simple Lie group is discrete (note Section 6.5 and Corollary 7.14).

3. Lie's Theorem on Solvability

We now describe how a solvable Lie group or Lie algebra of endomorphisms can be represented by triangular matrices. To do this we must compute characteristic roots so we consider real Lie groups or algebras as acting on complex vector spaces.

Definitions 10.15 (a) Let K be a field of characteristic 0 and let V be a finite-dimensional vector space over K. Let $T \in \operatorname{End}_K(V)$ and $\lambda \in K$. Then set

$$V_\lambda = \{X \in V\colon TX = \lambda X\}$$

and

$$V(\lambda) = \{X \in V\colon (T - \lambda I)^n X = 0 \text{ for some } n \in N\}$$

where N is the set of natural numbers (which are greater than 0). If $V_\lambda \neq \{0\}$, then λ is called a **characteristic value** or **eigenvalue** of T and $0 \neq X \in V_\lambda$ is called an **eigenvector** or **characteristic vector** of T with characteristic value λ. If $V(\lambda) \neq \{0\}$, then λ is called a **weight** of T and $V(\lambda)$ a **weight space** and $0 \neq X \in V(\lambda)$ is called a **weight vector** of T.

A characteristic value or a weight λ of T is a solution of the equation $\det(Ix - T) = 0$ and if all the solutions to this (characteristic) equation are in K, then we say that the **characteristic values or weights are in** K; recall the definition of a split endomorphism in Section 9.2.

(b) Let $N \subset \operatorname{End}_K(V)$, let $f\colon N \to K$ be a function, and set

$$V_f = \{X \in V\colon \text{for all } T \in N,\ TX = f(T)X\}$$

and

$$V(f) = \{X \in V\colon \text{for all } T \in N, \text{ there exists } n > 0 \text{ with } (T - f(T)I)^n X = 0\}.$$

If $V_f \neq \{0\}$, then f is called a **characteristic function** on N and $0 \neq X \in V_f$ is called a **characteristic vector** of N for the characteristic function f. Similarly one defines a **weight function, weight space,** and **weight vector** in case $V(f) \neq \{0\}$.

Thus these functions on N assign to each T in N a characteristic root $f(T)$ of T. Of course, in actual computations, the characteristic roots discussed above might be in the algebraic closure of K.

With these definitions and results on canonical forms of endomorphisms [Jacobson, 1953, Vol. II; Lang, 1965] we state the following:

Proposition 10.16 Let V be a finite-dimensional vector space over K and let $T \in \mathrm{End}_K(V)$ have its (distinct) weights $\lambda_1, \ldots, \lambda_m$ in K. Then the weight spaces $V(\lambda_i)$ are T-invariant and $V = V(\lambda_1) + \cdots + V(\lambda_m)$ (direct sum).

REMARK (1) This direct sum decomposition will be generalized in the next chapter to a direct sum decomposition of weight spaces of a nilpotent Lie group or Lie algebra.

Exercise (1) Let V be a finite-dimensional vector space over K, let g be a Lie subalgebra of $gl(V)$, and let $f: g \to K$ be a characteristic function on g. Show f is a linear transformation.

Proposition 10.17 Let V be a finite-dimensional vector space over R, let G be a Lie subgroup of $GL(V)$, and let $f: G \to R$ be a characteristic function with $f(G) \subset R^* = R - \{0\}$. Then regarding R^* as a multiplicative Lie group, the map $f: G \to R^*$ is an analytic homomorphism of Lie groups. f is frequently called a **character** of G.

PROOF Let $S, T \in G$. Then for $0 \neq X \in V_f$ we have $SX = f(S)X$ and $TX = f(T)X$. Thus

$$STX = Sf(T)X = f(T)SX = f(T)f(S)X.$$

However since $(ST)X = f(ST)X$ this gives $f(ST) = f(S)f(T)$ so that $f: G \to R^*$ is a homomorphism. To see that f is analytic, let $X_1, \ldots, X_m$ be a basis of V so that X_1 is a characteristic vector of G for the characteristic function f. Noting that the mappings $r: G \to V: S \to S(X_1)$ and $s: V \to R: \sum_{i=1}^m \lambda_i X_i \to \lambda_1$ are analytic, so is the map $f = s \circ r: G \to R^*$.

Analogous to Lemma 7.15 we have the following result:

Lemma 10.18 Let V be a finite-dimensional vector space over R and let G be a real connected Lie group which is a subgroup of $GL(V)$ and has real Lie algebra g. Let W be a subspace of V.

(a) W is invariant under the action of G if and only if W is invariant under the action of g.

(b) For $A \in g$, the vector $X \in V$ is a characteristic vector of A with characteristic value λ if and only if X is characteristic vector of the subgroup $\{\exp tA : t \in R\}$ for the characteristic function f: $\exp tA \to e^{t\lambda}$.

Exercise (2) Prove results analogous to Proposition 10.17 and Lemma 10.18 when we take V to be a finite-dimensional vector space over C and let G be a real Lie group which is a subgroup of $GL(V, C)$. For example, consider the real matrix Lie group $G = GL(n, R)$ as acting on C^n and regard G as a subgroup of $GL(n, C)$.

The following result or some of its equivalent consequences is known as "Lie's theorem on solvability." We follow the work of Tits [1965] for the group proof.

Theorem 10.19 (Lie's theorem) Let V be a finite-dimensional vector space over C and let G be a real connected solvable Lie group which is a subgroup of $GL(V, C)$. Then there exists a nonzero characteristic vector of G for some characteristic function.

PROOF We shall prove the results by induction on the dimension of G. First, if G is one dimensional and $0 \neq A \in g$ which is the Lie algebra of G, then since $g \subset gl(V)$ we see that A has a nonzero characteristic vector $X \in V$. However, by Lemma 10.18 and exercise (2), X is also a characteristic vector of G. Next assume G is of dimension n and assume as an induction hypothesis that we have shown the result for all such groups of smaller dimension. Now since G is connected and solvable, G has a connected solvable normal subgroup H of dimension $n-1$ [exercise (5), Section 10.2]. Thus by the induction hypothesis we can conclude there is a characteristic function f: $H \to C^* = C - \{0\}$ and analogous to Proposition 10.17 we have f is continuous.

We shall now show that the subspace $V_f = \{X \in V : SX = f(S)X$ for all $S \in H\}$ is invariant under G. Thus let $X \in V_f$, $S \in H$, and $T \in G$. Then

$$S(T(X)) = (ST)(X) = T(T^{-1}ST)(X) = f(T^{-1}ST)T(X) \qquad (*)$$

using $T^{-1}ST \in H$. Thus the number $f(T^{-1}ST)$ is a characteristic value of S with characteristic vector $T(X)$ and also the function $k : G \to C^* : T \to f(T^{-1}ST)$ is continuous. However, since G is connected and the set of characteristic values of S is discrete, the image $k(G)$ consists of a single point. (This uses the characterization: the topological space M is connected if and only if M is mapped continuously into a discrete space implies the image of M consists of a single point.) Thus we have $k(T) = k(I) = f(S)$. Using this in (*) we have for any $X \in V_f$, $T \in G$, and $S \in H$ that $S(T(X)) = f(S)T(X)$ which shows by the definition of V_f that V_f is invariant under the action of G.

Next by Lemma 10.18 we have that V_f is invariant under the action of $g \subset gl(V)$. Therefore if $A \in g$ and $A \notin h$ which is the Lie algebra of H, then, since the subspace V_f is invariant under the linear transformation A, there is a characteristic vector $0 \neq X \in V_f$ for A. Thus since $g/h = RA + h$ (using the hypothesis that dim H is $n - 1$) we see that X is a characteristic vector for g. For let $B = aA + bC \in g$ with $C \in h$ and for $AX = \lambda X \in V_f$ and $CX = \mu X$ [using Lemma 10.18(b) applied to h and H] we have

$$BX = (aA + bC)X = (a\lambda + b\mu)X.$$

Thus by Lemma 10.18, X is a characteristic vector for G.

Definition 10.20 Let V be an m-dimensional vector space over the field K. Then a sequence of subspaces $\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_m = V$ such that dim $V_i = i$ for $i = 1, \ldots, m$ is called a **flag** in V. Let $G \subset GL(V)$ be a Lie group. Then the flag is G-invariant if for every $T \in G$ we have $T(V_i) \subset V_i$ for $i = 1, \ldots, m$. Similarly for a Lie algebra g of endomorphisms, we define a g-invariant flag.

Proposition 10.21 Let V be an n-dimensional vector space over C and let G be a real connected Lie group which is a subgroup of $GL(V, C)$. Then the following are equivalent.

(a) The group G is solvable.

(b) There exists a flag which is G-invariant.

(c) There is a basis of V such that the matrices for the elements in G can be put simultaneously into triangular form. (The matrices might have complex entries).

PROOF Assume G is solvable. Then to show (b) we use induction on the dimension of V. From Lie's theorem there is a one-dimensional subspace W of V which is invariant under G. Therefore an element $T \in G$ induces a nonsingular linear map

$$\bar{T}: V/W \to V/W : x + W \to Tx + W$$

and the map $G \to GL(V/W, C) : T \to \bar{T}$ is an analytic homomorphism. Thus the image $\bar{G} = \{\bar{T} \in GL(V/W, C) : T \in G\}$ is a real connected solvable Lie group which is a subgroup of $GL(V/W, C)$ and by the induction hypothesis there exists a flag in V/W which is invariant under $\bar{G}$

$$\{\bar{0}\} \subset \bar{V}_2 \subset \bar{V}_3 \subset \cdots \subset \bar{V} = V/W.$$

Now let $\pi : V \to V/W$, let $V_i = \pi^{-1}(\bar{V}_i)$, and set $V_1 = W = \pi^{-1}(\{\bar{0}\})$. Then $\dim V_i = i$ and $\{0\} \subset V_1 \subset V_2 \subset \cdots \subset V_n = V$ is a flag which is invariant under G.

Next to show (b) implies (c) we choose a basis of V from the corresponding flag as follows. Let $V_1 = \{X_1\}$. Then since $TV_1 \subset V_1$ for all $T \in G$ we have $TX_1 = a_{11}(T)X_1$. Next let $V_2 = \{X_1, X_2\}$ where X_1 and X_2 are independent using $\dim V_2/V_1 = 1$. Then since $TV_2 \subset V_2$ we have $TX_2 = a_{12}(T)X_1 + a_{22}(T)X_2$ for all $T \in G$. Continuing in this manner we can choose a basis of V so that any $T \in G$ has a matrix of the form

$$\begin{bmatrix} a_{11}(T) & a_{12}(T) & \cdots & a_{1n}(T) \\ & a_{22}(T) & & \\ & & \ddots & \\ 0 & & & a_{nn}(T) \end{bmatrix}$$

with $0 \neq a_{11}(T) \cdots a_{nn}(T) = \det T$.

Finally to show (c) implies (a) let G be represented by the group triangular matrices as above. Let G_1 be the normal subgroup of triangular matrices of the form

$$\begin{bmatrix} 1 & & * & \\ 0 & 1 & & \\ & & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

with 1's on the diagonal. Let G_2 be the normal subgroup of G_1 of the form

$$\begin{bmatrix} 1 & 0 & & * & \\ & 1 & 0 & & \\ & & 1 & & \\ & & & \ddots & 0 \\ 0 & & & & 1 \end{bmatrix}$$

with 1's on the diagonal and 0's on the next superdiagonal. Continuing this way we obtain the sequence

$$G \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_n = \{I\}$$

with G_i/G_{i+1} commutative.

The preceding results on Lie groups can be translated into results on Lie algebras via the exp mapping or directly as follows. This proof involves some computations we shall see again in Chapter 11.

Theorem 10.22 (Lie's theorem) Let P be the algebraic closure of the field K and let V be a nonzero finite-dimensional vector space over P. Let g

be a solvable Lie algebra over K and let ρ be a homomorphism of g into $gl(V, P)$. Then there exists a vector $0 \neq X \in V$ which is a characteristic vector for all the members of $\rho(g)$ for some characteristic function.

PROOF We prove this by induction on the dimension of g. For $\dim g = 1$, the theorem follows from results on canonical forms (Proposition 10.16). We assume the results hold for Lie algebras of dimension less than $\dim g$. From Theorem 10.7(d) we can find an ideal h in g so that $\dim g/h = 1$. By Corollary 10.8 we have h is solvable so that by the induction assumption there exists a characteristic function $f: h \to P$ so that for all $S \in h$

$$\rho(S)X = f(S)X.$$

From $\dim g/h = 1$ we can find $T \in g$ so that $T \notin h$. Thus $g = KT + h$. Let W be the subspace of V spanned by all the vectors

$$X_1 = X \qquad \text{and} \qquad X_{k+1} = \rho(T)^k X$$

for $k = 1, 2, \ldots$. Note that W is $\rho(T)$-invariant subspace of V.

We shall now show: For all $S \in h$, $\rho(S)Y = f(S)Y$ for all $Y \in W$; that is, W is $\rho(h)$-invariant and furthermore $\rho(S) = f(S)I$ on W.

We first prove by induction that for all $S \in h$ and $k = 1, 2, \ldots$

$$\rho(S)X_k = f(S)X_k + a_{k-1}X_{k-1} + \cdots + a_1 X_1 \tag{$*$}$$

where $a_j = a_j(S)$ are in P. By the choice of $X_1 = X$ the result holds for $k = 1$. Assuming $(*)$ for k, we have

$$\begin{aligned}
\rho(S)X_{k+1} &= \rho(S)\rho(T)X_k, \qquad \text{definition of } X_{k+1} \\
&= \rho([ST])X_k + \rho(T)\rho(S)X_k \\
&= \rho([ST])X_k + \rho(T)\big(f(S)X_k + a_{k-1}X_{k-1} + \cdots + a_1X_1\big) \\
&= f(S)\rho(T)X_k + b_kX_k + \cdots + b_1X_1 \\
&= f(S)X_{k+1} + b_kX_k + \cdots + b_1X_1
\end{aligned}$$

using $[ST] \in h$ and the induction assumption.

We next prove $\rho(S)Y = f(S)Y$ for all $Y \in W$. From $(*)$ and the definition of X_k we first observe that W is $\rho(g)$-invariant. Next note that from the above, the restriction $\rho(S) \mid W$ has matrix

$$\begin{bmatrix} f(S) & & * \\ & \ddots & \\ 0 & & f(S) \end{bmatrix}$$

so that $\operatorname{tr} \rho(S) \mid W = f(S) \dim W$ for $S \in h$. Next note $\rho(S)$ and $\rho(T)$ map W into W so that $\rho([ST]) = \rho(S)\rho(T) - \rho(T)\rho(S)$ as endomorphisms of W.

Thus since $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for endomorphisms, we see $0 = \operatorname{tr} \rho([ST]) = f([ST]) \dim W$. Since $\dim W > 1$ this gives $f([ST]) = 0$. Thus

$$\begin{aligned} \rho(S)X_{k+1} &= \rho(S)\rho(T)X_k \\ &= \rho([ST])X_k + \rho(T)\rho(S)X_k \\ &= f([ST])X_k + f(S)\rho(T)X_k = f(S)X_{k+1} \end{aligned}$$

that is, $\rho(S)Y = f(S)Y$ for all $Y \in W$.

Since W is $\rho(T)$-invariant and P is algebraically closed we see that $\rho(T)$ has a characteristic vector $A \in W$: $\rho(T)A = tA$. Also $\rho(S)A = f(S)A$ for all $S \in h$ and since $g = KT + h$ we have for any $Z = aT + S$ that $\rho(Z)A = a\rho(T)A + \rho(S)A = (at + f(S))A$. Thus A is a characteristic vector of $\rho(g)$ and $F : aT + S \to at + f(S)$ defines the corresponding characteristic function.

The formalities in the proof of Proposition 10.21 yield the following:

Proposition 10.23 Let P be the algebraic closure of the field K and let V be a nonzero finite-dimensional vector space over P. Let g be a Lie algebra over K and let ρ be a homomorphism of g into $gl(V, P)$. Then the following are equivalent.

(a) The Lie algebra $\rho(g)$ is solvable,

(b) There is a flag in V which is invariant under $\rho(g)$.

(c) There is a basis of V such that the matrices for the endomorphisms in $\rho(g)$ can be put simultaneously into triangular form. (The matrices might have entries from P.)

These results apply when we take the field K to be algebraically closed itself. Thus $K = P$ and we obtain the following:

Proposition 10.24 Let g be a Lie algebra over the algebraically closed field K. Then g is solvable if and only if there exists a flag in g

$$\{0\} \subset g_1 \subset g_2 \subset \cdots \subset g_n = g$$

such that each g_i is an ideal of g.

PROOF Assume g is solvable. Then since $g \to \operatorname{ad}(g) : X \to \operatorname{ad} X$ is a homomorphism of Lie algebras over K, we see that $\operatorname{ad}(g)$ is a solvable Lie algebra of endomorphisms acting on the vector space g. By Proposition 10.23(b) there is a flag $\{0\} \subset g_1 \subset \cdots \subset g_n = g$ which is invariant under $\operatorname{ad}(g)$; that is, each g_i is an ideal of g.

Conversely, assuming such a flag exists we see that $\operatorname{ad}(g)$ is solvable; using (b) implies (a) in Proposition 10.23. However, $\operatorname{ad} : g \to \operatorname{ad}(g)$ is a homomorphism so that $\operatorname{ad}(g) \cong g/\ker(\operatorname{ad})$. Since $\ker(\operatorname{ad})$ is the center of g which is solvable and since $g/\ker(\operatorname{ad})$ is solvable, we have by Corollary 10.8 that g is solvable.

CHAPTER 11

NILPOTENT LIE GROUPS AND ALGEBRAS

We continue the concepts given in the preceding chapter and call a Lie group nilpotent if it is nilpotent as an abstract group. Then we discuss nilpotent Lie algebras and obtain the result that a connected Lie group is nilpotent if and only if its Lie algebra is nilpotent. In the last section we consider the vector space decomposition which yields the Jordan canonical form for an endomorphism and extend this decomposition to a nilpotent group of automorphisms.

1. Nilpotent Lie Groups

We now give a variation of the results on solvable groups using some of the notation of the preceding chapter.

Definition 11.1 Let G be an abstract group.

(a) Let $C^0G = G$ and let $C^{n+1}G = (G, C^nG)$. Then $C^nG \rhd C^{n+1}G$ and we have the **descending central series**

$$G = C^0G \rhd C^1G \rhd C^2G \rhd \cdots.$$

(b) Let $C_0 G = \{e\}$ and let $C_n G = \pi^{-1}(Z(G/C_{n-1}G))$, where $Z(G/C_{n-1}G)$ is the center of $G/C_{n-1}G$ noting by induction $C_n G \lhd C_{n+1}G \lhd G$ and where

$\pi : G \to G/C_{n-1}G$ is the corresponding projection map. Thus we have the **ascending central series**

$$\{e\} = C_0 G \lhd C_1 G \lhd C_2 G \lhd \cdots.$$

Theorem 11.2 Let G be an abstract group. Then the following are equivalent.

(a) There is a series of normal subgroups of G

$$G = G_0 \rhd G_1 \rhd \cdots \rhd G_s = \{e\}$$

such that $(G, G_n) \subset G_{n+1}$ for $n = 0, \ldots, s-1$.

(b) There exists a positive integer p such that

$$G \rhd C^1 G \rhd \cdots \rhd C^p G = \{e\}.$$

(c) There exists a positive integer q such that

$$\{e\} \lhd C_1 G \lhd \cdots \lhd C_q G = G.$$

PROOF Assume there is a series as given in (a). Then by induction we have $G_n \supset C^n G$. Thus $C^s G = \{e\}$. Conversely if (b) holds, then we automatically have a series satisfying (a).

Next we have (a) implies (c), for if we have a series as in (a), then we shall show by induction $G_{s-n} \subset C_n G$ so that for $n = s$ we obtain $C_s G = G$. Thus $\{e\} = G_s \subset C_0 G = \{e\}$ and assume $G_{s-i} \subset C_i G$ Then

$$(G/C_i G, G_{s-i-1}/C_i G) \subset G_{s-i}/C_i G \subset C_i G/C_i G = \{\bar{e}\}$$

using the induction hypothesis for the second inclusion; that is, $(G, G_{s-i-1}) \subset C_i G$. Thus if $\pi : G \to G/C_i G$ is the projection, we see that $G_{s-i-1} \subset \pi^{-1}(Z(G/C_i G)) = C_{i+1}G$, using the definition of $C_{i+1}G$.

Conversely to see (c) implies (a), we first note that

$$(G, C_i G)/C_{i-1}G \subset (G/C_{i-1}G, C_i G/C_{i-1}G) = \{\bar{e}\}$$

using $C_i G = \pi^{-1}(Z(G/C_{i-1}G))$, where $\pi : G \to G/C_{i-1}G$. Thus $(G, C_i G) \subset C_{i-1}G$ so that for $C_q G = G = G_0$, $C_{q-1}G = G_1, \ldots, C_{q-m}G = G_m$, etc., we see that the series in (c) yields the series in (a).

Definition 11.3 An abstract group G is **nilpotent** if it satisfies any one of the conditions of Theorem 11.2.

REMARKS (1) Note that nilpotency involves a descending series using commutators of the terms of the series with the group, whereas solvability involves a descending series using commutators of the terms of the series with itself.

(2) Subgroups, quotient groups, and finite direct products of nilpotent groups are nilpotent. The proofs run as expected. For example, if G_i are groups with $C^{n_i}G_i = \{e_i\}$ for $i = 1, \ldots, m$ and if $G = G_1 \times \cdots \times G_m$, then $C^nG = \{e\}$, where $n = \max\{n_1, \ldots, n_m\}$.

Example (1) Let V be an m-dimensional vector space over K and let $G \subset GL(V)$ be the set of triangular $m \times m$ matrices of the form

$$\begin{bmatrix} a & & & * \\ & a & & \\ & & \ddots & \\ 0 & & & a \end{bmatrix}$$

for $a \neq 0$ in K. Then G is nilpotent group, for let G_1 be the subgroup of matrices of the form

$$\begin{bmatrix} 1 & 0 & & & * \\ & 1 & & & \\ & & \ddots & & \\ & & & \ddots & 0 \\ 0 & & & & 1 \end{bmatrix}$$

and let G_2 be the subgroup of matrices of the form

$$\begin{bmatrix} 1 & 0 & 0 & & * \\ & 1 & 0 & & \\ & & \ddots & & 0 \\ & & & \ddots & 0 \\ 0 & & & & 1 \end{bmatrix}$$

Then $G \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_{m-1} = \{I\}$ and $(G, G_n) \subset G_{n+1}$.

Theorem 11.4 Let G be a Lie group. Then the following are equivalent.

(a) As an abstract group G is nilpotent.

(b) There exists a series $G = G_0 \rhd G_1 \rhd \cdots \rhd G_s = \{e\}$ where each G_n is a closed normal Lie subgroup of G and $(G, G_n) \subset G_{n+1}$.

(c) If $\bar{C}^0G = G$ and $\bar{C}^{n+1}G = \overline{(G, \bar{C}^nG)}$, then there exists a positive integer p such that $G \rhd \bar{C}^1G \rhd \cdots \rhd \bar{C}^pG = \{e\}$.

PROOF Showing (c) if and only if (b) is similar to Theorem 11.2; (b) implies (a) is also clear. Next assume (a). Then there is a series of normal subgroups $G \rhd G_1 \rhd \cdots \rhd G_s = \{e\}$ with $(G, G_n) \subset G_{n+1}$. Consequently we obtain $\bar{G}_s = \{e\}$, using G as Hausdorff, and $G \rhd \bar{G}_1 \rhd \cdots \rhd \bar{G}_s = \{e\}$ with $(G, \bar{G}_n) \subset \bar{G}_{n+1}$ which proves (b).

Definition 11.5 A Lie group G is a **nilpotent Lie group** if it is nilpotent as an abstract group.

Exercises (1) Let G be a Lie group. Show that $C^pG = \{e\}$ if and only if $\bar{C}^pG = \{e\}$ if and only if $C_pG = G$.

(2) Show that if G is a nilpotent Lie group and H is a closed normal subgroup of G, then G/H is a nilpotent Lie group.

(3) Let G be a Lie group and H a nilpotent subgroup. Then show its closure $\bar{H}$ is nilpotent.

2. Nilpotent Lie Algebras

Let G be a Lie group with Lie algebra g. We shall now define the notion of a nilpotent Lie algebra so that if G is connected, then G is a nilpotent Lie group if and only if g is a nilpotent Lie algebra.

Definition 11.6 (a) Let g be a Lie algebra over a field K and let $C^0g = g$ and $C^{n+1}g = [g\ C^ng]$. Thus we see that

$$C^1g = [gg], \ldots, C^kg = (\operatorname{ad} g)^k(g), \ldots$$

are ideals of g and we obtain the **descending central series** $g = C^0g \rhd C^1g \rhd \cdots$.

(b) Set $C_0g = \{0\}$ and $C_{n+1}g = \pi^{-1}(Z(g/C_ng))$, where by induction $C_ng \lhd g$ and $\pi : g \to g/C_ng$ is the Lie algebra homomorphism and $Z(g/C_ng)$ is the center of the Lie algebra g/C_ng. Thus we see that $C_0g = \{0\}$, $C_1g = Z(g)$, etc. are ideals of g and we obtain the **ascending central series** $\{0\} = C_0g \lhd C_1g \lhd \cdots$.

Theorem 11.7 Let g be a Lie algebra over K. Then the following are equivalent.

(a) There exists a sequence $g = g_0 \rhd g_1 \rhd \cdots \rhd g_s = \{0\}$ where all the g_n are ideals of g such that $[gg_n] \subset g_{n+1}$.

(b) There exists a positive integer p such that $g = C^0g \rhd C^1g \rhd \cdots \rhd C^pg = \{0\}$.

(c) There exists a positive integer q such that $\{0\} = C_0g \lhd C_1g \lhd \cdots \lhd C_qg = g$.

(d) There exists a positive integer r such that for all $X_1, \ldots, X_r \in g$ we have $\operatorname{ad} X_1 \circ \operatorname{ad} X_2 \circ \cdots \circ \operatorname{ad} X_r = 0$.

PROOF The equivalence of (a)–(c) are similar to Theorem 11.2. To see (b) if and only if (d) just use the fact that $C^k g$ is generated by the elements $(\operatorname{ad} X_1 \circ \operatorname{ad} X_2 \circ \cdots \circ \operatorname{ad} X_k)Y$ for any $X_1, \ldots, X_k$, $Y \in g$; see Definition 11.6.

Definition 11.8 A Lie algebra g is called **nilpotent** if it satisfies any one of the conditions of Theorem 11.7.

REMARKS (1) Subalgebras, quotient algebras, and finite direct sums of nilpotent Lie algebras are again nilpotent.

(2) A nilpotent Lie algebra is solvable, for by induction we obtain $C^n g \supset g^{(n+1)}$.

Exercises (1) What can be said about the following: Let g be a Lie algebra with h an ideal of g so that g/h is nilpotent and h is nilpotent. Then g is nilpotent?

(2) Let g denote the set of $m \times m$ matrices with elements in K of the form

$$\begin{bmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{mm} \end{bmatrix}$$

where $a_{11} = a_{22} = \cdots = a_{mm}$. Then g is a nilpotent Lie algebra, for let g_1 be the subalgebra of matrices of the form

$$\begin{bmatrix} 0 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}$$

and let g_2 be the subalgebra of matrices of the form

$$\begin{bmatrix} 0 & 0 & & & * \\ & 0 & \cdot & & \\ & & \ddots & \ddots & \\ & & & \ddots & 0 \\ 0 & & & & 0 \end{bmatrix}, \quad \text{etc.}$$

Show that $g \rhd g_1 \rhd g_2 \rhd \cdots \rhd g_m = \{0\}$ and $[gg_n] \subset g_{n+1}$.

Proposition 11.9 Let g be a Lie algebra over K. Then g is solvable if and only if $[gg]$ is nilpotent.

PROOF Suppose $[gg]$ is nilpotent. Then $[gg]$ is solvable. Also since $g/[gg]$ is a commutative Lie algebra, it is solvable. Thus by Corollary 10.8, g is solvable.

Conversely, first let P be the algebraic closure of K and let g be a solvable Lie algebra over P contained in $gl(V, P)$, where V is a finite-dimensional vector space over P. Then by the results following Lie's theorem (Proposition 10.23) there is a basis of V so that the matrices of g have triangular form

$$\begin{bmatrix} a_{11} & & & * \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{mm} \end{bmatrix}$$

and consequently the matrices for elements in $[gg]$ have the form

$$\begin{bmatrix} 0 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}.$$

Thus by exercise (2), $[gg]$ is a nilpotent Lie algebra.

Next if g is an arbitrary solvable Lie algebra over P, then $\mathrm{ad}(g)$ is solvable and therefore $[\mathrm{ad}(g), \mathrm{ad}(g)] = \mathrm{ad}([gg])$ is nilpotent. However, since $\mathrm{ad} : g \to \mathrm{ad}(g)$ is a Lie algebra homomorphism with $\ker(\mathrm{ad}) = Z(g)$ we see that $\bar{g} = g/Z(g) \cong \mathrm{ad}(g)$. Therefore $\bar{g}^{(2)} = [\bar{g}\bar{g}] \cong \mathrm{ad}([gg])$ is nilpotent. Consequently, there exists a positive integer p such that

$$\{\bar{0}\} = C^{p+1}\bar{g}^{(2)} = C^{p+1}g^{(2)}/Z(g).$$

Thus $C^{p+1}g^{(2)} \subset Z(g)$ so that $C^{p+2}g^{(2)} = \{0\}$; that is, $g^{(2)} = [gg]$ is nilpotent.

Finally, if g is a Lie algebra over K, we let $h = P \otimes_k g$ be the tensor product of the algebras P and g over K as in Section 9.1. Then h is a Lie algebra over P and a straightforward computation shows that if g is solvable, then h is solvable. Thus since $[gg] \subset [hh]$ we use the results of the preceding paragraph to conclude $[gg]$ is nilpotent.

Exercise (3) If g is a nilpotent Lie algebra over R, then show that its complexification g_C is also nilpotent.

Theorem 11.10 (Engel's theorem) Let V be a nonzero finite-dimensional vector space over the field K and let g be a Lie subalgebra of $gl(V)$ which consists of nilpotent linear transformations (that is, $A^n = 0$ for some n). Then there exists a nonzero vector $X \in V$ such that for all $A \in g$, we have $AX = 0$.

PROOF First we shall show that $A \in g$ being a nilpotent linear transformation implies $\mathrm{ad}_g A$ is a nilpotent linear transformation acting on g. Thus since $gl(V) = \mathrm{End}(V)$ as sets, we can define the endomorphisms

$$R(A) : gl(V) \to gl(V) : Z \to ZA \qquad \text{and} \qquad L(A) : gl(V) \to gl(V) : Z \to AZ$$

and see $(\mathrm{ad}\, A)(Z) = AZ - ZA = (L(A) - R(A))(Z)$ in $gl(V)$. Also noting that $L(A)R(A) = R(A)L(A)$ we have by the binomial theorem for any integer $k \geq 0$,

$$\begin{aligned}(\mathrm{ad}\, A)^k Z &= [L(A) - R(A)]^k Z \\ &= \sum_{i=0}^{k} (-1)^i \binom{k}{i} A^{k-i} Z A^i.\end{aligned}$$

However, since $A \in g$ is nilpotent, all the factors A^{k-i} or A^i are 0 for suitably large k or i. Thus $\mathrm{ad}_g A$ is nilpotent.

Next we shall use induction on $m = \dim g$ to prove the result. For $m = 1$ we have $g = KA$ where A is a nonzero nilpotent linear transformation. Thus since there exists $X \in V$ with $AX = 0$, the same result holds for every $B = bA \in g$. Now assume as an induction hypothesis that the result holds for Lie algebras of dimension less than m and let h be a proper subalgebra of g of maximum dimension. Then by the results of the above paragraph, $\mathrm{ad}_g A$ is a nilpotent endomorphism on g for all $A \in h$. Thus since $\mathrm{ad}\, A : h \to h$ we see $\mathrm{ad}\, A$ induces a nilpotent endomorphism $\bar{A}$ on the vector space $\bar{g} = g/h$. Furthermore the set $\bar{h} = \{\bar{A} : A \in h\}$ is a subalgebra of $gl(\bar{g})$ which consists of nilpotent endomorphisms and $\dim \bar{h} < m$.

By the induction hypothesis with $V = \bar{g}$ we can conclude that there exists $\bar{B} \neq \bar{0}$ in $\bar{g}$ such that for all $\bar{A} \in \bar{h}$ we have $\bar{A}\bar{B} = \bar{0}$; that is, there exists $B \in g$ with $B \notin h$ and $[h, B] \subset h$. Thus the subspace $h + KB$ of g is a subalgebra which contains h. However, by the maximal choice of h we have $h + KB = g$.

Finally let $W = \{Z \in V : AZ = 0 \text{ for all } A \in h\}$. Then by the above induction hypothesis $W \neq \{0\}$. Furthermore for $A \in h$ and $B \in g$ as above we have, since $[A, B] \in h$, that for any $Z \in W$,

$$A(BZ) = (AB)Z = [A, B]Z + (BA)Z = 0.$$

Thus by the definition of W we obtain $BW \subset W$. However, since $B \in g$ is nilpotent on V we have B is nilpotent on W. Consequently there exists $0 \neq X \in W$ with $BX = 0$ and since $g = h + KB$ we see this X has the desired property.

Corollary 11.11 Let V be a finite-dimensional vector space over K and let $g \subset gl(V)$ be a Lie algebra of nilpotent endomorphisms of V.

(a) There exists a basis of V such that the matrices of the endomorphisms in g relative to this basis have the form

$$\begin{bmatrix} 0 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}.$$

(b) g is a nilpotent Lie algebra of endomorphisms.

(c) The associative algebra g^* generated by the endomorphisms of g is a nilpotent associative algebra; that is, there exists a positive integer r such that for any endomorphisms $A_1, \ldots, A_r \in g^*$ we have $A_1 A_2 \cdots A_r = 0$.

Proof (a) Let $X_1 \in V$ be such that $AX_1 = 0$ for all $A \in g$. If the subspace $V_1 = KX_1 \neq V$, then each $A \in g$ induces a nilpotent endomorphism $\bar{A}$ on the nonzero vector space $\bar{V} = V/V_1$. Thus we can find $\bar{X}_2 = X_2 + V_1 \neq \bar{0}$ in $\bar{V}$ such that $\bar{A}\bar{X}_2 = \bar{0}$ for all $A \in g$; that is, there exists $X_2 \in V$ and $X_2 \notin V_1$ with

$$AX_2 = a_{21}(A)X_1 + 0X_2$$

for all $A \in g$. Continuing by induction we obtain a basis $X_1, \ldots, X_m$ of V such that for all $A \in g$,

$$AX_1 = 0 \qquad \text{and} \qquad AX_n \equiv 0 \mod (X_1, \ldots, X_{n-1})$$

where $(X_1, \ldots, X_{n-1})$ denotes the subspace spanned by these vectors. Thus the matrix for A has 0's on and below the diagonal.

Part (b) follows from (a) and exercise (2), while (c) follows from (a) and matrix multiplication.

Corollary 11.12 Let g be an abstract Lie algebra over K. Then g is a nilpotent Lie algebra if and only if for all $X \in g$ we have $\operatorname{ad}_g X$ is a nilpotent endomorphism on g.

Proof If g is nilpotent, then from Theorem 11.7(d) we have ad X is nilpotent. Conversely, if each ad X is nilpotent, then by Corollary 11.11, $(\operatorname{ad} g)^*$ is a nilpotent associative algebra. Thus there exists a positive integer p with $\{0\} = (\operatorname{ad} g)^p g = C^p g$; that is, g is nilpotent.

Theorem 11.13 Let G be a connected real Lie group with Lie algebra g. Then G is a nilpotent Lie group if and only if g is a nilpotent Lie algebra.

PROOF First assume that G is nilpotent and let $G = G_0 \rhd G_1 \rhd \cdots \rhd G_n = \{e\}$ be a series of closed normal subgroups of G such that $(G, G_k) \subset G_{k+1}$. Consequently we have the corresponding series $g = g_0 \rhd g_1 \rhd \cdots \rhd g_n = \{0\}$ of ideals of g. Next $(G, G_k) \subset G_{k+1}$ implies $[gg_k] \subset g_{k+1}$, for let $X \in g$, $Y \in g_k$. Then for t near $0 \in R$ we have from Theorem 5.16(c) that

$$(\exp tX, \exp tY) = \exp\bigl(t^2[XY] + o(t^3)\bigr)$$

is in G_{k+1}. However, from the characterization of the Lie algebra of G_{k+1} in Theorem 6.9, this implies $t^2[XY] + o(t^3) \in g_{k+1}$ which yields $[XY] \in g_{k+1}$; that is, $[gg_k] \subset g_{k+1}$.

We now sketch the main parts of the proof of the converse and leave the details as exercises. First, since g is a nilpotent Lie algebra, we see that ad g is a nilpotent Lie algebra of endomorphisms (with index of nilpotency N). Thus for any $Z \in g$, ad Z is nilpotent. Consequently in the expansion of Theorem 5.18

$$\exp X \cdot \exp Y = \exp F(X, Y)$$

for X, Y in g near the origin $0 \in g$, we have that the Campbell–Hausdorff formula

$$F(X, Y) = X + Y + \tfrac{1}{2}[XY] + \cdots$$

is of finite length since $(\operatorname{ad} X)^N = (\operatorname{ad} Y)^N = 0$.

Secondly, from the chain of ideals

$$g \rhd g_1 \rhd g_2 \rhd \cdots \rhd g_n = \{0\},$$

where $[gg_k] \subset g_{k+1}$, we obtain for the connected subgroup G_k generated by $\exp g_k$ the chain

$$G \rhd G_1 \rhd G_2 \rhd \cdots \rhd G_n = \{e\}.$$

Finally, for $X \in g$, $Y \in g_k$ near enough the origin 0, we have for $x = \exp X$, $y = \exp Y$

$$xyx^{-1}y^{-1} = \exp([XY] + \cdots) = \exp P(X, Y)$$

where $P(X, Y)$ is a finite sum of commutators, using the first part of the proof. Now each commutator term in $P(X, Y)$ contains $Y \in g_k$. However, g_k is an ideal of g so that $[gg_k] \subset g_{k+1}$ and therefore $P(X, Y) \in g_{k+1}$. Thus $xyx^{-1}y^{-1} \in \exp g_{k+1} \subset G_{k+1}$ and by induction on the length of products of elements G and G_k we obtain $(G, G_k) \subset G_{k+1}$.

Exercise (4) Show that each commutator term in $P(X, Y)$ contains $Y \in g_k$. Also complete the induction.

(5) Let g be a finite-dimensional Lie algebra over K.

(i) If h_1 and h_2 are ideals of g which are nilpotent Lie subalgebras, show that $h_1 + h_2$ is a nilpotent ideal of g.

(ii) Show that a maximal nilpotent ideal n of g exists; n is called the **nilpotent radical** of g. See the work of Bourbaki [1960] and Jacobson [1962] for more results of this nature.

3. Nilpotent Lie Algebras of Endomorphisms

We shall now generalize the process of finding the Jordan canonical form matrix of an endomorphism to the process of decomposing a vector space into weight spaces relative to a nilpotent Lie algebra of endomorphisms; that is, finding simultaneously "Jordan forms" for a nilpotent Lie algebra of endomorphisms. Recall from Section 9.2 that a Lie algebra over K is split if all the characteristic roots of ad X are in K for all $X \in g$.

Theorem 11.14 Let V be a finite-dimensional vector space over K, and let g be a split nilpotent Lie subalgebra of $gl(V)$.

(a) There exists a direct sum decomposition

$$V = V(\phi_1) + \cdots + V(\phi_m),$$

where $V(\phi_k) = \{X \in V : \text{for all } T \in g,\ (T - \phi_k(T)I)^p X = 0\}$ are g-invariant weight spaces for g for $k = 1, \ldots, m$.

(b) There exists a basis of V so that the matrices of the endomorphisms in g relative to this basis all have the block form

$$\begin{bmatrix} \begin{bmatrix} \phi_1(T) & & * \\ & \ddots & \\ 0 & & \phi_1(T) \end{bmatrix} & & 0 & \\ & \begin{bmatrix} \phi_2(T) & & * \\ & \ddots & \\ 0 & & \phi_2(T) \end{bmatrix} & & \\ 0 & & \ddots & \end{bmatrix}.$$

(c) The functions $\phi_k : g \to K$ are linear; that is, $\phi_k \in g^*$. Furthermore $\phi_k([g, g]) = \{0\}$.

PROOF We break the proof into several parts. First we have the following formula for an associative algebra A. Let $s, t \in A$, and let $s^{(0)} = s$, $s^{(1)} = ts - st = (\text{ad } t)s$, and $s^{(k)} = (\text{ad } t)^k s$. Then we obtain by induction for $k = 1, 2, \ldots$

$$\begin{aligned} t^k s &= \sum_{i=0}^{k} \binom{k}{i} s^{(i)} t^{k-i} \\ &= st^k + \binom{k}{l} s^{(1)} t^{k-1} + \cdots + s^{(k)}. \end{aligned} \tag{$*$}$$

Next we have the following result.

Lemma 11.15 Let V be a finite-dimensional vector space over K, and let g be a split nilpotent Lie subalgebra of $gl(V)$. Let $T \in g$, $\lambda \in K$ and let

$$V(\lambda) = \{X \in V : (T - \lambda I)^n X = 0 \text{ for some } n\}$$

be a weight space for T. Then $V(\lambda)$ is a g-invariant subspace of V.

PROOF We noted in Section 10.3 that $V(\lambda)$ is a subspace. Since g is a nilpotent Lie algebra, the Lie algebra $h = g + KI$ is also nilpotent where I is the identity endomorphism. Therefore, by Corollary 11.12, $[\text{ad}(T - \lambda I)]^N = 0$ for some fixed N. Now for $S \in g \subset \text{End}(V)$ let $S^{(1)} = [\text{ad}(T - \lambda I)]S$, $S^{(2)} = [\text{ad}(T - \lambda I)]^2 S$, etc. Then for $X \in V(\lambda)$ with $(T - \lambda I)^m X = 0$, we see by choosing $k = N + m$ and using $(*)$ that

$$(T - \lambda I)^k (SX) = \sum_{i=0}^{k} \binom{k}{i} S^{(i)} (T - \lambda I)^{k-i} X = 0$$

noting $S^{(N)} = [\text{ad}(T - \lambda I)]^N S = 0$. Thus by definition of $V(\lambda)$ we see $SX \in V(\lambda)$; that is, $V(\lambda)$ is g-invariant.

PROOF OF THEOREM 11.14 (continued) To prove part (a), we use induction on the dimension m of V. If $m = 1$, then every $T \in g$ has a characteristic root so that $TX = \lambda(T)X$ for $V = KX$. This yields the result in this case. For $m > 1$ we let $T \in g$, and by Proposition 10.16 we have the direct sum $V = V(\lambda_1) + \cdots + V(\lambda_n)$ where the $V(\lambda_i)$ are weight spaces for T. By Lemma 11.15 the $V(\lambda_i)$ are g-invariant, and consequently g restricts to a nilpotent Lie algebra of endomorphisms on each $V(\lambda_i)$. Thus we conclude the proof by induction, because we can assume T has at least two distinct characteristic roots (why?) so that the dimension of $V(\lambda_i)$ is less than the dimension of V. Now since V is finite dimensional, we see that there are only finitely many distinct weights ϕ_i of g.

To construct the basis of V which gives the matrix in (b), it suffices to find a basis for each $V(\phi_i)$ which gives the corresponding block matrix. Thus let $V(\phi)$ be a typical weight space as in (a). Then there is a nonzero $X \in V(\phi)$ such that $TX = \phi(T)X$. To see this, we use Lie's theorem (Theorem 10.23) replacing algebraic closure by "split" as follows. Since g is nilpotent on $V(\phi)$, it is solvable on $V(\phi)$. Thus there is a nonzero $X \in V(\phi)$ and a characteristic function F so that for all $T \in g$, $TX = F(T)X$. Therefore

$$[F(T) - \phi(T)]X = [T - \phi(T)I]X,$$

and by induction

$$[F(T) - \phi(T)]^k X = [T - \phi(T)I]^k X = 0$$

for k large enough, remembering $X \in V(\phi)$. Thus $F(T) = \phi(T)$; that is, $\phi(T)$ is the only characteristic root.

Thus, for $X_1 = X$ as above, the one-dimensional subspace $W = KX_1$ is g-invariant. Set $\bar{V} = V(\phi)/W$. Then g induces a nilpotent Lie algebra of endomorphisms $\bar{g}$ by $\bar{T}\bar{X} = \overline{TX}$. From this the characteristic roots of $\bar{T}$ are $\phi(T)$, and $\bar{V}$ is a weight space of dimension less than the dimension of V. By induction we can find a basis $\bar{X}_2, \ldots, \bar{X}_m$ of $\bar{V}$ so that

$$\begin{aligned}
\bar{T}\bar{X}_2 &= \phi(T)\bar{X}_2 \\
\bar{T}\bar{X}_3 &= a_{23}(\bar{T})\bar{X}_2 + \phi(T)\bar{X}_3 \\
&\vdots \\
\bar{T}\bar{X}_m &= \sum_{j=2}^{m-1} a_{jm}(\bar{T})\bar{X}_j + \phi(T)\bar{X}_m .
\end{aligned}$$

Thus for $\bar{X}_i = X_i + W$ and $W = KX_1$, we can find a basis $X_1, \ldots, X_m$ of $V(\phi)$ so that

$$\begin{aligned}
TX_1 &= \phi(T)X_1 \\
TX_2 &= a_{12}(T)X_1 + \phi(T)X_2 \\
&\vdots \\
TX_m &= \sum_{j=1}^{m-1} a_{jm}(T)X_j + \phi(T)X_m .
\end{aligned}$$

Thus we have the desired basis for $V(\phi)$.

For part (c), we show a weight ϕ is a linear functional as follows. As in (b), let $0 \neq X \in V(\phi)$ be such that for all $T \in g$, $TX = \phi(T)X$. Then for $S, T \in g$ we have $S + T \in g$ and

$$\phi(S + T)X = (S + T)X = SX + TX = [\phi(S) + \phi(T)]X$$

so that $\phi(S + T) = \phi(S) + \phi(T)$. Similarly $\phi(aS) = a\phi(S)$ for $a \in K$, and also $\phi([S, T])X = [S, T]X = (ST)X - (TS)X = \phi(S)\phi(T)X - \phi(T)\phi(S)X = 0.$

Thus since the elements of $[g, g]$ are of the form $\sum [S, T]$, this implies $\phi([g, g]) = 0$.

Exercises (1) State results analogous to those in Theorem 11.14 for nilpotent Lie groups. [Recall if $f: G \to R^*$ is an analytic character (Proposition 10.17), then $Tf(e): g \to R$ is an algebra homomorphism.]

(2) (i) Let $\mathscr{D}$ be a nilpotent Lie algebra of derivations of a nonassociative algebra A over an algebraically closed field. Decompose $A = A(\phi_1) + \cdots + A(\phi_k)$ into weight spaces relative to $\mathscr{D}$. Show $A(\phi_i)A(\phi_j) = \{0\}$ if $\phi_i + \phi_j$ is not a weight, and $A(\phi_i)A(\phi_j) \subset A(\phi_i + \phi_j)$ if $\phi_i + \phi_j$ is a weight.

(ii) Let g be a Lie algebra over an algebraically closed field and let h be a nilpotent subalgebra of g. Then what can be said about a weight space decomposition of g relative to ad h?

(iii) Let g be a Lie algebra over an algebraically closed field, and let D be a nonsingular derivation of g. Show that g is a nilpotent Lie algebra.

CHAPTER 12

SEMISIMPLE LIE GROUPS AND ALGEBRAS

We have previously defined a Lie group G to be semisimple in case its Lie algebra g is semisimple. Consequently in this chapter we discuss generalities on semisimple Lie algebras over a field K of characteristic 0 and in the remaining chapters investigate them in more detail. First we consider a nonassociative algebra A with a certain invariant bilinear form $f : A \times A \to K$ and show that if this form is nondegenerate, then A is a direct sum of ideals which are simple algebras. We apply this to the case where A is associative and the form $f(X, Y) = \text{trace}\, L(X)L(Y)$ and discuss the semisimplicity of A.

With the associative algebra as a model, we prove results due to Cartan which lead to the fact that a Lie algebra g over K is semisimple if and only if the form $f(X, Y) = \text{trace ad}\, X \,\text{ad}\, Y$ is nondegenerate. As immediate corollaries we prove that a nonzero ideal of a semisimple Lie algebra is also semisimple and that a derivation of a semisimple Lie algebra is inner. We eventually use this to show that for a large class of "semisimple" nonassociative algebras every derivation is inner.

Next we come to the very important result that a Lie algebra g is semisimple if and only if every g-module is completely reducible. Using this we show that if a Lie algebra g of endomorphisms acts in a completely reducible manner on a vector space V, then $g = c \oplus g'$, where c is the center of g and $g' = [g, g]$ is semisimple or $\{0\}$. As applications of these results we discuss the nilpotent radical of a Lie algebra and the tensor product of completely reducible g-modules.

1. Invariant Bilinear Forms

We shall show in this section how certain bilinear forms can be used to decompose nonassociative algebras into a direct sum of ideals which are simple algebras. In particular, this will yield for Lie and associative algebras that they are semisimple; that is, the algebras contain no proper solvable ideals.

Definitions 12.1 (a) Let g be a Lie algebra over the field K of characteristic 0 and let $\rho : g \to gl(V)$ be a finite-dimensional representation of g where V is a finite-dimensional vector space over K. Then the map

$$\tau : g \times g \to K : (X, Y) \to \text{trace } \rho(X) \circ \rho(Y)$$

is a symmetric bilinear form on g called the **trace form** for g relative to ρ. In particular, for $V = g$ and $\rho = \text{ad}$ the corresponding trace form is called the **Killing form** and we shall frequently denote this form by Kill(X, Y).

(b) Let f be a bilinear form on the vector space V, let $G \subset GL(V)$ be a Lie group, and let $g \subset gl(V)$ be a Lie algebra. Then f is called **G-invariant** if for all $X, Y \in V$ and $A \in G$ we have

$$f(AX, AY) = f(X, Y).$$

Similarly f is **g-invariant** if for all $X, Y \in V$ and $D \in g$ we have

$$f(DX, Y) + f(X, DY) = 0.$$

Note Corollary 12.6 for the relationship between the G-invariance and the g-invariance of f when g is the Lie algebra of G.

Proposition 12.2 Let g be a Lie algebra and let ρ be a finite-dimensional representation of g in V. Then the trace form $\tau(X, Y) = \text{trace } \rho(X)\rho(Y)$ is ad(g)-invariant.

PROOF First recall for endomorphisms S and T that trace ST = trace TS. So let $X, Y, Z \in g$, then for $D = ad(X)$ we have

$$\begin{aligned}
\tau(DY, Z) &= \tau([XY], Z) \\
&= \text{trace } \rho([XY])\rho(Z) \\
&= \text{trace } [\rho(X), \rho(Y)]\rho(Z) \\
&= \text{trace } \rho(X)\rho(Y)\rho(Z) - \text{trace } \rho(Y)\rho(X)\rho(Z) \\
&= \text{trace } \rho(Y)\rho(Z)\rho(X) - \text{trace } \rho(Y)\rho(X)\rho(Z) \\
&= \text{trace } \rho(Y)\rho([ZX]) = -\tau(Y, DZ).
\end{aligned}$$

Corollary 12.3 (a) The Killing form of g is ad(g)-invariant; that is, for $X, Y, Z \in g$ we have

$$\mathrm{Kill}([XY], Z) = \mathrm{Kill}(X, [YZ]).$$

(b) The Killing form of g is Der(g)-invariant, where Der(g) is the derivation algebra of g.

PROOF To see (b) note that if D is a derivation of g, then $\mathrm{ad}(DX) = [D, \mathrm{ad}\, X]$. Thus as in the calculations in Proposition 12.2 we have

$$\begin{aligned}\mathrm{Kill}(DX, Y) &= \mathrm{trace\ ad}(DX)\,\mathrm{ad}(Y)\\ &= \mathrm{trace}\,(D \cdot \mathrm{ad}\, X - \mathrm{ad}\, X \cdot D)\,\mathrm{ad}\, Y\\ &= \mathrm{trace\ ad}\, X(\mathrm{ad}\, YD - D\,\mathrm{ad}\, Y)\\ &= -\mathrm{trace\ ad}\, X\,\mathrm{ad}\, DY = -\mathrm{Kill}(X, DY).\end{aligned}$$

Definition 12.4 Let A be a nonassociative algebra over a field K. A symmetric bilinear form $f: A \times A \to K$ is called **invariant** or **associative** if for all $X, Y, Z \in A$ we have

$$f(XY, Z) = f(X, YZ).$$

Proposition 12.5 Let V be a finite-dimensional vector space over R and let $A \in \mathrm{End}(V)$ and let f be a bilinear form on V. Then the following are equivalent.

(a) For all $X, Y \in V$ we have $f(AX, Y) + f(X, AY) = 0$.

(b) For all $X, Y \in V$ and $t \in R$ we have $f((\exp tA)X, (\exp tA)Y) = f(X, Y)$.

PROOF If we assume (b), then we obtain (a) by using the product rule for differentiation as in Section 1.2. Conversely, assuming (a), we obtain (b) by showing $\phi(t) = f((\exp tA)X, (\exp tA)Y)$ and $\psi(t) = f(X, Y)$ are both solutions to the differential equation $dz/dt = 0$ with $z(0) = f(X, Y)$. Clearly $\psi(t)$ is a solution. Next using the product rule we obtain

$$\phi'(t) = f((A \exp tA)X, (\exp tA)Y) + f((\exp tA)X, (A \exp tA)Y) = 0,$$

where we use (a) replacing X [respectively Y] by $(\exp tA)X$ [respectively $(\exp tA)Y$] to obtain the last equality. Thus by uniqueness of solutions $\phi(t) = \psi(t)$.

Corollary 12.6 Let $G \subset GL(V)$ be a connected Lie group of automorphisms of V over R and let G have Lie algebra $g \subset gl(V)$. Let f be a bilinear form on V. Then f is G-invariant if and only if f is g-invariant.

Exercises (1) Let g be a Lie algebra over R and let f be a bilinear form defined on g. Show f is invariant under the identity component of $\mathrm{Aut}(g)$ if and only if f is invariant under $\mathrm{Der}(g)$.

(2) The Killing form of a Lie algebra g over K is $\mathrm{Aut}(g)$-invariant. [Note for $f \in \mathrm{Aut}(g)$ that $f \circ \mathrm{ad}\, X \circ f^{-1} = \mathrm{ad}(fX)$.]

Definition 12.7 Let W be a subspace of V and let f be a symmetric or skew-symmetric bilinear form on V. Then

$$W^{\perp} = \{X \in V : \text{for all } Y \in W, f(X, Y) = 0\}$$

is the **orthogonal complement** of W relative to f. Then f is **nondegenerate** if $V^{\perp} = \{0\}$.

From results in algebra [Jacobson, 1953, Vol. II; Lang, 1965] we have the following:

Proposition 12.8 Let f be a nondegenerate symmetric bilinear form on V and let W be a subspace of V such that $W \cap W^{\perp} = \{0\}$; that is, W is nonisotropic. Then $V = W + W^{\perp}$ as a subspace direct sum.

Lemma 12.9 Let A be a finite-dimensional nonassociative algebra over K with a symmetric nondegenerate invariant form $f : A \times A \to K$ and let B be an ideal in A. Then:

(a) $B^{\perp}$ (relative to f) is an ideal in A.
(b) $B \cap B^{\perp}$ is an ideal with $(B \cap B^{\perp})^2 = \{0\}$.

PROOF For any $X \in A$, $Y \in B$, and $Z \in B^{\perp}$ we use $AB \subset B$ to obtain

$$f(ZX, Y) = f(Z, XY) \in f(B^{\perp}, B) = \{0\}$$

so that $B^{\perp}A \subset B^{\perp}$. Similarly

$$f(XZ, Y) = f(Y, XZ) = f(YX, Z) \in f(B, B^{\perp}) = \{0\}$$

so that $AB^{\perp} \subset B^{\perp}$, which yields $B^{\perp}$ is an ideal of A.

To show (b) we use $B \cap B^{\perp}$ is in both B and $B^{\perp}$ to obtain

$$f((B \cap B^{\perp})^2, A) = f(B \cap B^{\perp}, (B \cap B^{\perp})A).$$

However, $B \cap B^{\perp} \subset B^{\perp}$ and $(B \cap B^{\perp})A \subset BA \subset B$ so that using $f(B^{\perp}, B) = \{0\}$ we obtain

$$f(B \cap B^{\perp}, (B \cap B^{\perp})A) = \{0\}.$$

Thus $(B \cap B^{\perp})^2 \subset A^{\perp} = \{0\}$ using f is nondegenerate.

Following Jacobson's proof [1962] we now prove a structure theorem, due to Dieudonné, for algebras with a nondegenerate invariant form. Recall that a nonassociative algebra is **simple** if $A^2 \neq \{0\}$ and A has no proper ideals.

Theorem 12.10 Let A be a finite-dimensional nonassociative algebra over K with a symmetric nondegenerate invariant form f. Furthermore assume A has no ideals B with $B^2 = \{0\}$. Then A is a direct sum of ideals which are simple subalgebras.

PROOF Since A is finite dimensional we let B be a nonzero minimal ideal of A. Then $B \cap B^{\perp}$ is an ideal of A such that $(B \cap B^{\perp})^2 = \{0\}$, using Lemma 12.9. Thus, using the hypothesis, we have $B \cap B^{\perp} = \{0\}$ so that by Proposition 12.8 we have $A = B \oplus B^{\perp}$ (direct sum of ideals of A) with $B^{\perp}$ an ideal of A and furthermore $BB^{\perp} \subset B \cap B^{\perp} = \{0\}$. From this B is a simple algebra for if C is a proper ideal of B, then since $CB^{\perp} = B^{\perp}C = \{0\}$ we see C is an ideal of A. This contradicts the minimal choice of B.

Next we see that the restriction $\bar{f}$ of f to $B^{\perp} \times B^{\perp}$ defines an invariant form on $B^{\perp}$ which is also nondegenerate, for if $f(B^{\perp}, X) = 0$ with $X \in B^{\perp}$, then since $A = B \oplus B^{\perp}$ we have $f(A, X) = \{0\}$ so that $X = 0$. Thus $B^{\perp}$ satisfies the same hypothesis as A and by induction on the dimension we can conclude $B^{\perp} = A_2 \oplus \cdots \oplus A_m$, where the A_k are ideals of $B^{\perp}$ (and therefore of A) which are simple subalgebras. Therefore with $A_1 = B$ we have the desired results $A = A_1 \oplus A_2 \oplus \cdots \oplus A_m$.

We also have the following uniqueness for the above decomposition.

Proposition 12.11 Let A be a finite-dimensional nonassociative algebra such that

$$A = A_1 \oplus \cdots \oplus A_m = B_1 \oplus \cdots \oplus B_n,$$

where the A_i and B_j are ideals of A which are simple subalgebras. Then $m = n$ and for each A_i there is a $B_{j(i)}$ with $A_i = B_{j(i)}$.

PROOF For each $j = 1, \ldots, n$ consider the ideal $A_1 \cap B_j$. If $A_1 \cap B_j = \{0\}$ for each $j = 1, \ldots, n$, then we have

$$A_1 B_j \subset A_1 \cap B_j = \{0\}$$

using both A_1 and B_j as ideals of A. Thus since $A = B_1 \oplus \cdots \oplus B_n$ we obtain $A_1 A = \{0\}$ so that $A_1{}^2 = \{0\}$ which contradicts the simplicity of A_1. Thus for some $j = j(1)$ we have the ideal $A_1 \cap B_{j(1)} \neq \{0\}$ so that

$$A_1 = A_1 \cap B_{j(1)} = B_{j(1)}$$

using the simplicity of the ideals A_1 and $B_{j(1)}$. Similarly there exists $j(2)$ with $A_2 = B_{j(2)}$ so that $A_1 \oplus A_2 = B_{j(1)} \oplus B_{j(2)}$. Thus we can conclude the proof using induction by noting every B_j is equal to some $A_{i(j)}$.

Exercise (3) Let g be a finite-dimensional Lie algebra over K with $g = g_1 \oplus g_2 \oplus \cdots \oplus g_n$ where the g_j are ideals of g which are simple subalgebras. Show $g = [gg]$ and for every ideal p of g there exists a unique ideal q of g so that $g = p \oplus q$ and $p \cap q = \{0\}$. Is a similar result valid for a general nonassociative algebra over K?

Example (1) Let A be a finite-dimensional associative algebra and $L(X) : A \to A : Y \to XY$ be the left multiplication by X. Let

$$f(X, Y) = \operatorname{trace} L(X)L(Y).$$

Then using $L(XY)Z = (XY)Z = X(YZ) = L(X)L(Y)Z$ we see

$$f(X, Y) = \operatorname{trace} \mathrm{L}(XY).$$

Using A is associative, this shows f is an invariant form on A.

Next if f is nondegenerate, we shall show there are no ideals B of A with $B^2 = \{0\}$; for suppose B is such an ideal and extend a basis of B to a basis of A as follows: $\{X_1, \ldots, X_r, X_{r+1}, \ldots, X_m\}$ is a basis of A where $X_1, \ldots, X_r \in B$. Then for $Z \in B$ and $X \in A$, $L(Z)$ and $L(X)$ have the following block matrices, respectively,

$$\begin{bmatrix} 0 & Z_{12} \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix}$$

which uses B as an ideal and $ZB = \{0\}$. Thus we obtain

$$f(Z, X) = \operatorname{trace} \begin{bmatrix} 0 & Z_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix} = 0,$$

and since we are assuming f is nondegenerate we have $Z = 0$; that is, $B = \{0\}$. This proves the following result.

Proposition 12.12 Let A be a finite-dimensional associative algebra over K and let $f(X, Y) = \operatorname{tr} L(X)L(Y)$ be nondegenerate. Then $A = A_1 \oplus \cdots \oplus A_n$ where each A_k is an ideal of A which is a simple subalgebra.

From Chapter 9 recall that a simple associative algebra is isomorphic to a complete matrix algebra over some division ring. We now sketch some relationships between the radical of a finite-dimensional associative algebra A and $f(X, Y) = \operatorname{tr} L(X)L(Y)$. First, the **radical** N of A is usually characterized as the maximal nilpotent ideal of A. However, using associativity we see N is also the maximal solvable ideal of A.

Lemma 12.13 Let $A \in \text{End}(V)$, where V is an m-dimensional vector space over a field K of characteristic 0. Then A is nilpotent if and only if trace $A^k = 0$ for $k = 1, 2, \ldots, m$.

PROOF From results on the characteristic polynomial $\det(A - xI)$ we have that trace $A = \sum_i \lambda_i$ where λ_i are the characteristic roots of A. Since the characteristic roots of A^k are $(\lambda_i)^k$ we also have trace $A^k = \sum_i (\lambda_i)^k$. First, if A is nilpotent, then extend A to an endomorphism of $\bar{V} = P \otimes V$ where P is the algebraic closure of K. Thus since A is nilpotent on V, it is nilpotent on $\bar{V}$ so all the characteristic roots are 0 and therefore trace $A^k = 0$ for $k = 1, 2, \ldots, m$.

For the converse we use the fact that

$$F(x) = \det(xI - A) = x^m - p_1 x^{m-1} + \cdots + (-1)^m p_m$$

where

$$\begin{aligned} p_1 &= \sum \lambda_i \\ p_2 &= \sum_{i>j} \lambda_i \lambda_j \\ &\vdots \\ p_n &= \lambda_1 \lambda_2 \ldots \lambda_m \end{aligned}$$

are the elementary symmetric polynomials in the characteristic roots $\lambda_i \in P$; that is, writing $F(x) = (x - \lambda_1) \cdots (x - \lambda_m)$ in $P[x]$. Next use the fact for $s_k = \sum_i (\lambda_i)^k$ for $k = 1, \ldots, m$ that we have the following relation proved by induction [Jacobson, 1953, Vol. I, p. 110].

$$0 = s_k - p_1 s_{k-1} + p_2 s_{k-2} + \cdots + (-1)^{k-1} p_{k-1} s_1 + (-1)^k p_k.$$

Thus using the hypothesis that $s_k = \text{trace } A^q = 0$ for $q = 1, \ldots, m$ we have $p_k = 0$ for $k = 1, \ldots, m$ so that $F(x) = x^m$. Thus $A^m = 0$.

Proposition 12.14 Let A be a finite-dimensional associative algebra over a field K of characteristic 0, let N be the radical of A, and let $f(X, Y) =$ trace $L(X)L(Y)$. Then

$$N = A^\perp = \{Z \in A : f(Z, X) = 0 \text{ for all } X \in A\}.$$

PROOF Let $Z \in N$. Then for any $Y \in A$ we have $L(Z)Y \in N$ and $[L(Z)]^{k+1} Y \in N^k = N \cdots N$, k-times. Thus since N is nilpotent, $L(Z)$ is nilpotent. However, for any $X \in A$ we have $L(Z)L(X) = L(ZX)$ is nilpotent since $ZX \in N$. Therefore

$$f(Z, X) = \text{trace } L(ZX) = 0$$

so that $N \subset A^\perp$.

Conversely, let $Z \in A^{\perp}$ and let $T = L(Z)^2 = L(Z^2)$. Then

$$\text{trace } T = f(Z, Z) = 0.$$

Next using associativity we have

$$T^2 = L(Z^2)L(Z^2) = L(Z^3)L(Z)$$

and by induction

$$T^k = L(Z^{2k-1})L(Z).$$

Thus since $Z \in A^{\perp}$ we obtain trace $T^k = 0$ for $k = 1, 2, \ldots$ which implies T is nilpotent. Thus $L(Z)$ is nilpotent. Therefore if $m = \dim A$, then the ideal $A^{\perp}$ is such that each $Z \in A^{\perp}$ satisfies $Z^{m+1} = 0$ and this implies $A^{\perp}$ is nilpotent using the exercise below. Since the radical N is the maximal nilpotent ideal, $A^{\perp} \subset N$.

Exercise (4) Let C be an ideal of the finite-dimensional associative algebra A so that for each $Z \in C$, there exists m with $Z^m = 0$. Then C is nilpotent; that is, there exists N so that $\{0\} = C^N = C \cdots C$, N-times.

Corollary 12.15 Let A be a finite-dimensional associative algebra over a field K of characteristic 0. If A has no proper nilpotent ideals, then $A = A_1 \oplus \cdots \oplus A_m$ where each A_k is an ideal which is a simple subalgebra.

Exercise (5) Let A be a finite-dimensional associative algebra over R. Show A is semisimple over R if and only if $C \otimes_R A$ is semisimple over C. (Possibly compare the degeneracy of the trace form on A with that on $C \otimes_R A$.) This result extends to a field K of characteristic 0 and using the results in Section 9.5 on complete reducibility show the following: V is a completely reducible A-module over K if and only if $P \otimes_K V$ is a completely reducible $P \otimes_K A$-module where P is the algebraic closure of K.

2. Cartan's Criteria

We now prove results for Lie algebras analogous to those on associative algebras involving bilinear forms and semisimplicity.

Theorem 12.16 (Cartan's criterion for solvability) Let V be a finite-dimensional vector space over a field K of characteristic 0 and let g be a Lie subalgebra of $gl(V)$. Then g is solvable if and only if trace$(XY) = 0$ for all $X \in g$ and $Y \in [gg]$.

Recall from Section 10.2 that a Lie algebra is semisimple if it has no proper solvable ideals.

Theorem 12.17 (Cartan's criterion for semisimplicity) Let g be a finite-dimensional Lie algebra over a field of characteristic 0. Then g is semisimple if and only if its Killing form $\text{Kill}(X, Y) = \text{trace ad } X \text{ ad } Y$ is nondegenerate.

For the proof of Cartan's criterion for solvability we use the methods of Bourbaki [1960], Serre [1965], and Tits [1965]. These differ from the methods of Hausner and Schwartz [1968] and Jacobson [1962] since the latter introduce the Cartan subalgebra first to obtain Theorem 12.16. However, in both cases a careful examination of certain rational-valued linear functionals is necessary. We shall need the following facts on linear algebra [Jacobson, 1953, Vol. II; Lang, 1965].

Definition 12.18 Let V be a finite-dimensional vector space over a field K of characteristic 0. Then $A \in \text{End}(V)$ is called a **semisimple endomorphism** if the associative algebra $K[A]$ generated by A is a semisimple algebra; note $I = A^0 \in K[A]$.

REMARKS (1) We have $A \in \text{End}(V)$ is semisimple if and only if its minimum polynomial $\mu(x) \in K[x]$ is a product of distinct prime polynomials. For suppose A is semisimple and its minimum polynomial $\mu(x) = p_1(x)^{k(1)} \cdots p_r(x)^{k(r)}$, where some $k(i) > 1$. Then the element $Z = p_1(A) \cdots p_r(A)$ is nilpotent since $Z^q = p_1(A)^q \cdots p_r(A)^q = 0$ for large enough q. Since $K[A]$ consists of polynomials in A and is commutative, we see that $K[A] \cdot Z$ is a nilpotent ideal in $K[A]$. This contradicts the semisimplicity of $K[A]$. Conversely, suppose the $p_k(x)$ are distinct and each $k(i) = 1$. Assume $K[A]$ is not semisimple so that some $Z = f(A)$ is nilpotent. Then $0 = Z^k = f(A)^k$ implies that the polynomial $f(x)^k$ is divisible by $\mu(x)$. Thus since all the $k(i) = 1$, $f(x)$ is divisible by $\mu(x)$; that is, $f(x) = g(x)\mu(x)$ so that $Z = f(A) = g(A)\mu(A) = 0$.

(2) Using the result that an associative algebra B with $B^2 \neq \{0\}$ is semisimple if and only if every B-module is completely reducible (Section 9.5) we see that the endomorphism A is semisimple if and only if for any $K[A]$-invariant subspace W of V, there exists a $K[A]$-invariant subspace W' so that $V = W + W'$ with $W \cap W' = \{0\}$.

(3) If K is algebraically closed, we can compute characteristic roots and vectors to obtain A as a semisimple endomorphism if there exists a basis

$X_1, \ldots, X_m$ of V so that $AX_i = \lambda_i X_i$ for some $\lambda_i \in K$; that is, A is diagonalizable. [This uses remark (1).]

Using Jordan canonical forms we have the following result:

Proposition 12.19 Let V be a finite-dimensional vector space over the algebraically closed field K of characteristic 0 and let $A \in \mathrm{End}(V)$. Then there exists a semisimple $S \in \mathrm{End}(V)$ and a nilpotent $N \in \mathrm{End}(V)$ such that

$$A = S + N \qquad \text{and} \qquad SN = NS.$$

Furthermore, S and N are uniquely determined by these conditions. There exist polynomials $s(x), n(x) \in K[x]$ without constant terms such that $S = s(A)$ and $N = n(A)$. Then S (respectively N) is called the **semisimple** (respectively **nilpotent) component** of A.

PROOF Let $F(x) = \det(xI - A) = (x - \lambda_1)^{k_1} \cdots (x - \lambda_r)^{k_r}$ be the factorization of the characteristic polynomial where the $\lambda_1, \ldots, \lambda_r$ are distinct. Then following the proof of Jacobson [1953, Vol. II, p. 130] the

$$\mu_i(x) = F(x)/(x - \lambda_i)^{k_i}$$

for $i = 1, \ldots, r$ are relative prime. Thus there exist $\phi_k(x) \in K[x]$ with

$$\sum_{i=1}^{r} \phi_i(x)\mu_i(x) = 1.$$

Substituting A into this expression we obtain

$$\sum \phi_i(A)\mu_i(A) = I$$

which gives the direct sum decomposition

$$V = V_1 + \cdots + V_r,$$

where the V_j are the A-invariant subspaces

$$\begin{aligned} V_j &= \phi_j(A)\mu_j(A)V \\ &= \{X \in V : (A - \lambda_j I)^{k_j} X = 0\}. \end{aligned}$$

Thus define S by $SX_j = \lambda_j X_j$ for $X_j \in V_j$ and set $N = A - S$. Then S is semisimple and N is nilpotent and furthermore, by construction, $[A, S] = 0$ so that $NS = SN$.

A straightforward computation shows that for the polynomial $s(x) = \sum \lambda_i \phi_i(x)\mu_i(x)$ we have $S = s(A)$ and for $n(x) = x - s(x)$ we have $N = n(A)$. We can furthermore assume $s(0) = 0$ as follows. If A is invertible, then the characteristic polynomial $F(x)$ has a constant term so that I is a polynomial in A without constant term. Thus if I appears in the expression $S = s(A)$, make the above substitution to obtain another polynomial expression for S in terms of A without constant terms. In this case we also see $n(0) = 0$. If A is not invertible, then $\ker A \neq \{0\}$ is invariant under N and $N|\ker(A)$ is also nilpotent. Thus there is a nonzero $X \in \ker(A)$ such that $NX = 0$. Thus since $N = n(A)$ and $AX = 0$, we see that the constant term of $n(x)$ is 0. Thus $n(0) = 0$ so that $s(0) = 0$ from the construction $n(x) = x - s(x)$.

Exercise (1) The representation $A = S + N$ with $SN = NS$ is unique; that is, if $A = S + N = S' + N'$, where S, S' are semisimple and N, N' nilpotent and $[S, N] = [S', N'] = 0$, then $S = S'$ and $N = N'$.

Example (1) Let $A \in gl(V)$, where V is finite-dimensional over an algebraically closed field K, and let $A = S + N$ be the decomposition into semisimple and nilpotent components. Then ad S and ad N are the semisimple and nilpotent components of ad A. Since $[S, N] = 0$ we see $0 = \operatorname{ad}[S, N] = [\operatorname{ad} S, \operatorname{ad} N]$ so that we must show ad N is nilpotent and ad S is semisimple.

First for any $X \in gl(V)$ we see

$$(\operatorname{ad} N)X = NX - XN = (L(N) - R(N))X,$$

where $L(N)$ and $R(N)$ are left and right multiplications in $\operatorname{End}(V)$ and $[L(N), R(N)] = 0$ using the associative law. Thus using the binomial expansion we obtain as before

$$\begin{aligned}(\operatorname{ad} N)^k X &= (L(N) - R(N))^k X \\ &= \sum_{i=0}^{k} (-1)^i C_{k,i} N^{k-i} X N^i.\end{aligned}$$

Thus if $N^q = 0$, ad N is nilpotent. Next since S is semisimple, there exists a basis $X_1, \ldots, X_m$ of V so that $SX_i = \lambda_i X_i$ with $\lambda_i \in K$. For the basis $\{E_{ij} : i, j = 1, \ldots, m\}$ of $gl(V)$ defined by $E_{ij} X_k = \delta_{jk} X_i$ we obtain the usual matrices and consequently the matrix computations give

$$(\operatorname{ad} S)E_{ij} = (\lambda_i - \lambda_j)E_{ij}.$$

Thus ad S is diagonalizable on $gl(V)$; that is, ad S is semisimple.

Lemma 12.20 Let V be a finite-dimensional vector space over an algebraically closed field K of characteristic 0. Let T and W be subspaces of $gl(V)$ with $W \supset T$ and let $P = \{A \in gl(V) : [A, W] \subset T\}$. Let $A \in P$ be such that

$$\operatorname{trace} AZ = 0$$

for all $Z \in P$. Then A is nilpotent.

PROOF Let $A = S + N$ be the decomposition of A into its semisimple and nilpotent components and let $X_1, \ldots, X_m$ be a basis of V so that $SX_i = \lambda_i X_i$ with $\lambda_i \in K$ and N is given by a nilpotent matrix. Now let L be the vector space spanned by the λ_i's over the rational field Q. Thus $L \subset K$ and to show A is nilpotent we shall show $L = \{0\}$ by proving $f(L) = \{0\}$ for every Q-linear functional $f \in L^*$.

Thus let $X_1, \ldots, X_m$ be the above basis of V and let $U \in \operatorname{End}(V)$ be the semisimple endomorphism defined by

$$UX_k = f(\lambda_k)X_k,$$

where f is any given linear functional in L^*. Let $\{E_{ij} : i, j = 1, \ldots, m\}$ be the basis of $gl(V)$ defined by the basis $X_1, \ldots, X_m$ as in the preceding example. Then from this example we can deduce

$$(\operatorname{ad} S)^k E_{ij} = (\lambda_i - \lambda_j)^k E_{ij} \qquad \text{for} \quad k = 0, 1, 2, \ldots$$

and

$$(\operatorname{ad} U)E_{ij} = (f(\lambda_i) - f(\lambda_j))E_{ij} = f(\lambda_i - \lambda_j)E_{ij},$$

using U as semisimple and f as linear.

Next by the interpolation formula we can find a polynomial $p(x) \in K[x]$ which goes through the finitely many points $0, f(\lambda_i - \lambda_j)$. Thus $p(x)$ has 0 constant term and $p(\lambda_i - \lambda_j) = f(\lambda_i - \lambda_j)$. Using this and the above formula for $(\operatorname{ad} S)^k E_{ij} = (\lambda_i - \lambda_j)^k E_{ij}$ we see

$$\begin{aligned}(\operatorname{ad} U)E_{ij} &= f(\lambda_i - \lambda_j)E_{ij} \\ &= p(\lambda_i - \lambda_j)E_{ij} = p(\operatorname{ad} S)E_{ij}\end{aligned}$$

so that $\operatorname{ad} U = p(\operatorname{ad} S)$. From the example, $\operatorname{ad} S = s(\operatorname{ad} A)$, where $s(x)$ is a polynomial without constant term, and since $(\operatorname{ad} A)W \subset T$ we see $(\operatorname{ad} U)W = [(p \circ s)(\operatorname{ad} A)]W \subset T$. Thus by the definition of P, we have $U \in P$.

Since $U \in P$ we have by hypothesis and the construction of U, S, and N relative to the same basis $X_1, \ldots, X_m$ of V, that

$$0 = \operatorname{trace} AU = \operatorname{trace} (S + N)U = \operatorname{trace} SU = \sum \lambda_i f(\lambda_i).$$

Therefore since $f(\lambda_i) \in Q$ and f is a Q-linear functional, we apply f to the above equation to obtain

$$0 = \sum f(\lambda_i)^2$$

which implies $f(\lambda_i) = 0$. Thus $f(L) = \{0\}$ and since f is arbitrary in L^* we have $L = \{0\}$, so that A is nilpotent.

We now use this to prove Cartan's criterion for solvability: If g is a subalgebra of $gl(V)$ where V is finite-dimensional over a field K of characteristic 0, then g is solvable if and only if trace $XY = 0$ for all $X \in g$ and $Y \in [g, g]$.

First assume the field K is algebraically closed and suppose trace $XY = 0$ for all $X \in g$ and $Y \in [g, g]$. Then we shall apply Lemma 12.20 with $W = g$ and $T = [g, g]$ to obtain

$$[g, g] \subset P = \{Z \in gl(V) : [Z, g] \subset [g, g]\}.$$

Thus for any X, $Y \in g$ and $Z \in P$ we have

$$\text{trace } [X, Y]Z = \text{trace } X[Y, Z] = 0,$$

where we use trace $XA = 0$ for all $X \in g$ and $A \in [g, g]$ for the second equality. Thus by the linearity of "trace" on $gl(V)$ we have trace $AZ = 0$ for all $A \in [g, g]$ and $Z \in P$. Therefore by Lemma 12.20, every $A \in [g, g]$ is a nilpotent endomorphism. Thus by Corollary 11.12 to Engel's theorem $[g, g]$ is a nilpotent Lie algebra. However since $g/[g, g]$ is Abelian and $[g, g]$ solvable, we have by Corollary 10.8 that g is solvable.

Conversely, if g is solvable, then by Lie's theorem (Proposition 10.23) we can find a basis which simultaneously puts the matrices for $[g, g]$ in triangular form with 0's on the diagonal. Thus for $X \in g$ and $Y \in [g, g]$,

$$\text{trace } XY = \text{trace} \begin{bmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix} \begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{bmatrix} = 0$$

which proves the result if K is algebraically closed.

Next if the field K is not algebraically closed, let P be its algebraic closure and let $\tilde{g} = P \otimes_K g$. If g is solvable, then from Section 10.2, $\tilde{g}$ is solvable. Thus trace $\bar{X}\bar{Y} = 0$ for all $\bar{X} \in \tilde{g}$ and $\bar{Y} \in [\tilde{g}, \tilde{g}]$ implies trace $XY = 0$ for all $X \in g$ and $Y \in [g, g]$.

Conversely, if trace $XY = 0$ for all $X \in g$ and $Y \in [g, g]$, then for any $\bar{X} = \sum w_i X_i \in \tilde{g}$ and $\bar{Y} = \sum u_{pq}[X_p, X_q] \in [\tilde{g}, \tilde{g}]$, where $w_i, u_{pq} \in P$ and $X_1, \ldots, X_m$ is a basis of g, we obtain

$$\text{trace } \bar{X}\bar{Y} = \sum w_i u_{pq} \text{ trace } X_i[X_p, X_q] = 0.$$

Thus $\tilde{g}$ is solvable and since $g \subset \tilde{g}$ we have that g is solvable.

Exercise (2) Using the fact that for A, $B \in \text{End}(V)$,

$$\begin{aligned} 2 \text{ trace } AB &= \text{trace}(AB + BA) \\ &= \text{trace}(A + B)^2 - \text{trace } A^2 - \text{trace } B^2 \end{aligned}$$

show the following result:

Corollary 12.21 If g is a Lie algebra over a field of characteristic 0, then g is solvable if and only if $\text{trace}(\text{ad } X)^2 = 0$ for all $X \in [gg]$.

We can now prove Cartan's criterion for semisimplicity: If g is a Lie algebra over a field of characteristic 0, then g is semisimple if and only if the Killing form of g is nondegenerate.

First assume g is semisimple and let $h = g^{\perp} = \{Z \in g : \text{Kill}(X, Z) = 0$ for all $X \in g\}$. Then h is an ideal of g. Furthermore we have for each $Z \in h$ and $Y \in [gg]$ that

$$\text{trace ad } Z \text{ ad } Y = \text{Kill}(Z, Y) = 0.$$

Therefore by Cartan's criterion for solvability we see ad h is a solvable subalgebra of $gl(g)$. However, ad h is isomorphic to h/Z where $Z = \ker \text{ad}$ is commutative. Thus since h/Z is solvable (since ad h is solvable) and Z is solvable, h is a solvable ideal. Therefore $h \subset \text{rad } g = \{0\}$ which shows the Killing form is nondegenerate.

Conversely, assume that the Killing form is nondegenerate. Then g has no proper ideals h with $[hh] = \{0\}$; for assume h is such an ideal and extend a basis of h to a basis of g. Thus as in example (1) in Section 12.1 for an associative algebra, we have for $Z \in h$ and $X \in g$ that

$$\text{ad } Z = \begin{bmatrix} 0 & Z_{12} \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \text{ad } X = \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix}$$

so that $\text{Kill}(Z, X) = 0$; that is, $h = \{0\}$. However, if $r = \text{rad } g \neq 0$, then in the series $r \rhd r^{(2)} \rhd \cdots \rhd r^{(n-1)} \rhd r^{(n)} = \{0\}$ we see $r^{(n-1)}$ is a nonzero ideal of g which satisfies $[r^{(n-1)}r^{(n-1)}] = 0$ (This uses that if h and k are ideals of a Lie algebra g, then the product $[hk]$ is an ideal of g). Thus this contradiction shows $r = \{0\}$.

Corollary 12.22 Let g be a semisimple Lie algebra over a field K of characteristic 0 and let ρ be a faithful (injective) representation of g in a finite-dimensional vector space V over K. Then the bilinear form $\tau(X, Y) = \text{trace } \rho(X)\rho(Y)$ is nondegenerate.

Proof Analogous to the preceding proof one sees that $h = \{Z \in g : \tau(Z, Y) = 0$ for all $Y \in g\}$ is an ideal of g such that $\rho(h)$ is solvable in $gl(V)$. However, since $\ker \rho = \{0\}$ we have h is solvable. Thus $h = \{0\}$ so that τ is nondegenerate.

Corollary 12.23 Let g be a finite-dimensional Lie algebra over a field K of characteristic 0. Then:

(a) g is semisimple if and only if $g = g_1 \oplus \cdots \oplus g_n$ where each g_k is an ideal of g which is a simple subalgebra;

(b) g is semisimple if and only if g has no ideals h with $[hh] = \{0\}$.

PROOF This follows from the fact that g is semisimple if and only if Kill(X, Y) is nondegenerate, and Theorem 12.10 which then expresses g as a direct sum of simple subalgebras which are ideals.

REMARKS (4) We shall now outline the proof that if g is a simple Lie algebra of endomorphisms over an algebraically closed field K of characteristic 0, then for all X, $Y \in g$

$$\text{Kill}(X, Y) = \lambda \text{ trace } XY$$

for some $\lambda \in K$.

Exercises (3) Let f be any nondegenerate invariant form on the simple Lie algebra g. Show that there exists $S \in GL(g)$ so that for all X, $Y \in g$

$$f(X, Y) = \text{Kill}(SX, Y) \qquad \text{and} \qquad \text{Kill}(SX, Y) = \text{Kill}(X, SY).$$

(4) Using the invariance and nondegeneracy of f and Kill, show [ad X, S] $= 0$ for all $X \in g$. Since g is simple, conclude by Schur's lemma [Proposition 9.14(a)] that $S = \lambda I$.

(5) (i) Let $f(X, Y) = \text{trace } XY$ and show f is an invariant form.

(ii) Use the fact that g is simple to show that $g^\perp$ (relative to f) is g or $\{0\}$.

(iii) Use Lemma 12.20 and (ii) to show that f is nondegenerate and conclude the desired formula.

(iv) What can be said about an invariant form of a semisimple Lie algebra?

(6) Let g be the Lie algebra of $n \times n$ matrices of trace 0 over the field K as above.

(i) Show g is simple as follows. With the usual matrix basis E_{ij} note that the E_{ij} for $i \neq j$ and $H_k = E_{kk} - E_{nn}$ for $k \leq n - 1$ form a basis of g. Relative to this basis of g find the multiplication table. Assume h is a nonzero proper ideal and let $Z = \sum a_i H_i + \sum z_{ij} E_{ij} \in h$. Carefully grind out products $[H_i, [H_j, Z]]$ and compare with Z to eventually show $h = g$. A better method will be given later.

(ii) Show Kill(X, Y) $= 2n$ trace XY for g as in (i).

(7) Let g be a semisimple Lie algebra and let h be a semisimple subalgebra.

(i) Show that $\rho : h \to \text{ad}(g) : X \to \text{ad } X$ is a faithful representation of h in g where ad is the adjoint in g. Thus g is an h-module.

(ii) Show that g has the direct sum decomposition $g = m + h$ where $m = h^\perp$, the orthogonal complement of h relative to the Killing form; note Corollary 12.22.

(iii) For the decomposition in (ii) define an anticommutative multiplication XY on m by $XY = [XY]_m$ which is the projection of $[XY]$ in g into m. Denote this algebra by (m, XY). For $X, Y \in m$ let $f(X, Y) = \operatorname{Kill}(X, Y)$ and show f is a nondegenerate invariant form on (m, XY). Sagle and Winter [1967] show that if g is simple, h is semisimple, and the multiplication $mm \neq \{0\}$, then (m, XY) is a simple algebra.

3. Ideals and Derivations of Semisimple Lie Algebras

If h is an ideal of semisimple Lie algebra g over a field of characteristic 0, then we shall show that h itself is a semisimple Lie algebra. In particular, this will yield the proposition that all the derivations of g map h into h and also that $\operatorname{Der}(g) = \operatorname{ad}(g)$; that is, the derivations of g are inner.

Theorem 12.24 Let h be an ideal of a Lie algebra g over a field of characteristic 0 and let Kill (respectively C) denote the Killing form of g (respectively h).

(a) If $X, Y \in h$, then $C(X, Y) = \operatorname{Kill}(X, Y)$.

(b) If g is semisimple and $A \in h$ is such that $C(A, Y) = 0$ for all $Y \in h$, then $A = 0$.

(c) If g is semisimple, then h is a semisimple Lie algebra.

PROOF (a) Since h is an ideal of g we can decompose the vector space $g = h + b$ (subspace direct sum) where $[hb] \subset h$ and $[hh] \subset h$. Thus choosing a basis from this decomposition we obtain for any $X \in h$,

$$\operatorname{ad}_g X = \begin{bmatrix} X_{11} & X_{12} \\ 0 & 0 \end{bmatrix}$$

where X_{11} is a matrix for $\operatorname{ad}_h X$. Thus for $X, Y \in h$ we have

$$\begin{aligned} \operatorname{Kill}(X, Y) &= \operatorname{trace} \begin{bmatrix} X_{11} & X_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{12} \\ 0 & 0 \end{bmatrix} \\ &= \operatorname{trace} X_{11} Y_{11} \\ &= \operatorname{trace} \operatorname{ad}_h X \operatorname{ad}_h Y = C(X, Y). \end{aligned}$$

(b) Let $A \in h$ be such that $C(A, Y) = 0$ for all $Y \in h$ and let $W \in g$. Then we use $[XW] \in h$ for any $X \in h$ to obtain

$$\operatorname{Kill}([AX], W) = \operatorname{Kill}(A, [XW]) = C(A, [XW]) = 0$$

using part (a) for the second equality. However since we are assuming g to be semisimple, we have by Cartan's criterion for semisimplicity (Theorem 12.17) that $[AX] = 0$ for all $X \in h$. Thus using $g = h + b$ we see

$$\mathrm{ad}_g A = \begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix}$$

and since

$$\mathrm{ad}_g W = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix}$$

we obtain

$$\mathrm{Kill}(A, W) = \mathrm{trace}\begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix} = 0.$$

Thus $A = 0$ using the nondegeneracy of the Killing form of g.

(*c*) Use part (b) and Cartan's criterion for semisimplicity.

Exercise (1) Let g be a semisimple Lie algebra over a field of characteristic 0 and let $g = g(1) \oplus \cdots \oplus g(n)$ be its decomposition into simple ideals. If h is an ideal of g, show $h = g(i_1) \oplus \cdots \oplus g(i_k)$ for suitable $i_1, \ldots, i_k$. In particular, this shows $[hh] = h$.

Corollary 12.25 Let h be an ideal of a semisimple Lie algebra g over a field of characteristic 0.

(a) If $g/h \neq \{\bar{0}\}$, then g/h is a semisimple Lie algebra.
(b) If D is a derivation of g, then $Dh \subset h$.

Proof Since g/h is not $\{\bar{0}\}$ we use $g = g(1) \oplus \cdots \oplus g(n)$ and $h = g(i_1) \oplus \cdots \oplus g(i_k)$ to see that g/h is isomorphic to the semisimple Lie algebra $g(j_1) \oplus \cdots \oplus g(j_s)$ where $j_t \notin \{i_1, \ldots, i_k\}$. To prove (b) we use $h = [hh]$ and $D[XY] = [DXY] + [XDY]$ to obtain

$$Dh = D[hh] \subset [Dhh] + [hDh] \subset h$$

since h is an ideal of g.

Theorem 12.26 If g is a semisimple Lie algebra over a field K of characteristic 0, then $\mathrm{Der}(g) = \mathrm{ad}(g)$.

Proof We follow the proof given by Jacobson [1962] where just the nondegeneracy of the Killing form is used. Thus let $D \in \mathrm{Der}(g)$ and let

$$\phi : g \to K : X \to \mathrm{trace}(\mathrm{ad}\, X \circ D).$$

Then ϕ is a linear functional on g and since the Killing form of g is nondegenerate, there exists $A \in g$ such that for all $X \in g$

$$\phi(X) = \text{Kill}(X, A).$$

(See the works of Lang [1965] and Jacobson [1953, Vol. II] for the representation of linear functionals in terms of a nondegenerate bilinear form.)

Let $E = D - \text{ad } A$. Then E is a derivation of g and for any $X \in g$ we have

$$\begin{aligned}\text{trace}(\text{ad } X \circ E) &= \text{trace}(\text{ad } X \circ D) - \text{trace ad } X \circ \text{ad } A \\ &= \phi(X) - \text{Kill}(X, A) = 0.\end{aligned}$$

Using this we have for any X, $Y \in g$ that

$$\begin{aligned}\text{Kill}(EX, Y) &= \text{trace ad}(EX) \text{ ad } Y \\ &= \text{trace}[E, \text{ad } X] \text{ ad } Y \\ &= \text{trace}(E \text{ ad } X \text{ ad } Y - \text{ad } X\, E \text{ ad } Y) \\ &= \text{trace}(E \text{ ad } X \text{ ad } Y - E \text{ ad } Y \text{ ad } X) \\ &= \text{trace } E[\text{ad } X, \text{ad } Y] \\ &= \text{trace } E \text{ ad}[XY] = 0\end{aligned}$$

where the second equality uses $\text{ad}(EX) = [E, \text{ad } X]$ for any derivation E and the fourth equality uses $\text{trace } PQ = \text{trace } QP$ for endomorphisms P and Q. Thus by the nondegeneracy of the Killing form we have $EX = 0$ so that $D = \text{ad } A$ is inner.

Exercise (2) According to Definition 7.7 what can be said about the automorphisms of a semisimple Lie algebra over R?

4. Complete Reducibility and Semisimplicity

In this section we shall discuss results for Lie algebras analogous to those for associative algebras concerning semisimplicity and complete reducibility. In particular, we shall prove Weyl's theorem which states that a Lie algebra g is semisimple if and only if every g-module is completely reducible. Then we apply this to find the structure of a Lie algebra of endomorphisms which acts in a completely reducible manner on a vector space. As before, let K denote a field of characteristic 0.

Recall the following from Section 9.5. Let V be a finite-dimensional g-module and let ρ be the corresponding representation. Then V or ρ is called

irreducible or **simple** if the g-module V has no proper g-submodules. Also V or ρ is called **completely reducible** or **semisimple** if the g-module V is a direct sum of irreducible g-submodules.

As for associative algebras, V is a completely reducible g-module if and only if for every g-submodule W of V, there exists a g-submodule W' so that $V = W + W'$ (subspace direct sum). For example, if g is a semisimple Lie algebra, then g is a completely reducible g-module. We now follow Bourbaki's proof [1960] of Weyl's theorem.

Lemma 12.27 Let g be a Lie algebra over the field K, let h be an ideal of g, and let ρ be a representation of g in the finite-dimensional vector space V. Assume that the bilinear form

$$\tau(X, Y) = \text{trace}\ \rho(X)\rho(Y) \qquad \text{for} \quad X, Y \in g$$

is nondegenerate when restricted to $h \times h$. Let $\{X_1, \ldots, X_n\}$ be a basis of h and let $\{X_1', \ldots, X_n'\} \subset h$ be a dual basis; that is, $\tau(X_i, X_j') = \delta_{ij}$.

(a) The element

$$C = \sum_{i=1}^{n} \rho(X_i) \circ \rho(X_i')$$

is an endomorphism of V which commutes with every endomorphism $\rho(A)$ for $A \in g$. We call C the **Casimir operator** (of h) corresponding to ρ.

(b) If V is an irreducible g-module, then the Casimir operator of g is an automorphism of V.

(c) trace $(C) = n$ $(= \dim h)$, where C is the Casimir operator of h.

Proof A straightforward calculation shows that the bilinear form τ is invariant. Thus for A, X, $Y \in g$ we have $\tau([AX], Y) + \tau(X, [AY]) = 0$. Now for $A \in g$ and X_i, $X_k' \in h$ as above set

$$[AX_i] = \sum_j a_{ji} X_j \qquad \text{and} \qquad [AX_k'] = \sum_p a'_{pk} X_p'$$

using $[gh] \subset h$. Then we obtain

$$a_{ji} = \tau([AX_i], X_j') = -\tau(X_i, [AX_j']) = -a'_{ij}.$$

Now for any $A \in g$ we have

$$\begin{aligned}
[\rho(A), C] &= [\rho(A), \textstyle\sum \rho(X_i)\rho(X_i')] \\
&= \sum_i ([\rho(A), \rho(X_i)]\rho(X_i') + \rho(X_i)[\rho(A), \rho(X_i')]) \\
&= \sum_i (\rho([AX_i])\rho(X_i') + \rho(X_i)\rho([AX_i'])) \\
&= \sum_{ij} a_{ji}\rho(X_j)\rho(X_i') + a'_{ji}\rho(X_i)\rho(X_j') = 0
\end{aligned}$$

using $a_{ji} = -a'_{ij}$. Thus the endomorphism C commutes with every endomorphism $\rho(A)$ for all $A \in g$. Next we have

$$\begin{aligned}\operatorname{trace}(C) &= \sum_i \operatorname{trace} \rho(X_i)\rho(X_i') \\ &= \sum_i \tau(X_i, X_i') = n\end{aligned}$$

using $\tau(X_i, X_j') = \delta_{ij}$ and $n = \dim h$. Thus $C \neq 0$. In particular, for the Casimir operator C of g we have from $[C, \rho(g)] = \{0\}$ that $\ker C$ is $\rho(g)$-invariant. Thus by Schur's lemma [Proposition 9.14(a)], C is an automorphism if V is g-irreducible.

Exercises (1) Show the Casimir operator is independent of the choice of basis.

(2) The following result on linear algebra is used in the next proof. Let W be a subspace of the vector space V and let $T \in \operatorname{End}(V)$ be such that $T: W \to W$ and $T \mid W = \lambda I$, where I is the identity endomorphism on W. Then $T \mid W$ is called a **homothetic** endomorphism (relative to W).

(i) Show that $M = \{T \in \operatorname{End}(V) : T(V) \subset W \text{ and } T \mid W \text{ is homothetic}\}$ is a subspace which contains the subspace $N = \{T \in M : T(W) = \{0\}\}$.

(ii) Decompose $V = W + P$ into a subspace direct sum and let $T \in M$ be as in (i). Relative to a basis chosen from this decomposition show T has a matrix of the form

$$\begin{bmatrix}\lambda I & t_{12} \\ 0 & 0\end{bmatrix}$$

for $\lambda \in K$. Thus show that $T = \lambda E + T'$, where $E^2 = E$ and $T' \in N$ is as in (i).

(iii) Use (ii) to show N is of codimension one in M; that is, $\dim M/N = 1$. Also show we can choose $EV = W$ and $P = \ker E$ in (ii). Thus $V = W + \ker E$ is a direct sum.

Lemma 12.28 Let g be a Lie algebra over the field K. Then the following are equivalent.

(a) Every representation of g in a finite-dimensional vector space is completely reducible.

(b) If ρ is a representation of g in a finite-dimensional vector space M and if N is a subspace of M of codimension 1 so that for every $X \in g$ we have $\rho(X)M \subset N$, then there exists a direct sum complement of N in M which is annihilated by $\rho(g)$.

PROOF First assume (a). Then for the subspace N of (b) we have $\rho(g)N \subset N$ so that N is actually a g-submodule. By the complete reducibility there exists a submodule N' with $M = N + N'$ (direct sum). Thus $\rho(g)N' \subset N'$ but also by hypothesis $\rho(g)N' \subset N$. Consequently $\rho(g)N' \subset N \cap N' = \{0\}$ which proves (b).

Next assume (b) and let σ be a representation of g in the finite-dimensional vector space V; that is, $\sigma(g)$ is a Lie subalgebra of $gl(V)$. Let W be $\sigma(g)$-submodule of V. Then we shall show W has a $\sigma(g)$-invariant direct complement in V. Let ν denote the adjoint representation of $gl(V)$ and define μ by

$$\mu = \nu \circ \sigma : g \to \mathrm{ad}(gl(V));$$

that is, for $X \in g$ we have $\mu(X) = \mathrm{ad}(\sigma(X))$ so that for $T \in gl(V)$ we have $\mu(X)T = [\sigma(X), T]$. Thus μ is a representation of g in the space $gl(V)$. Let

$$\begin{aligned} M &= \{T \in gl(V) : T(V) \subset W \text{ and } T \mid W \text{ is homothetic}\}, \\ N &= \{T \in M : T(W) = 0\}. \end{aligned}$$

Then from the preceding exercise we have N is a subspace of M such that $\dim M/N = 1$ and also $\mu(g)M \subset N$ as follows. Let $X \in g$, $T \in M$, and $A \in W$. Then using W is $\sigma(g)$-invariant

$$\begin{aligned} (\mu(X)T)A &= [\sigma(X), T]A \\ &= \sigma(X)(TA) - T\sigma(X)A \\ &= \sigma(X)(\lambda A) - \lambda(\sigma(X)A) = 0 \end{aligned}$$

so that $\mu(X)T \in N$.

Now from our asumption (b) we can find a direct sum complement N' of N in M which is annihilated by $\mu(g)$. Since $M = N + N'$ and $\dim N' = \dim M/N = 1$ we can assume $N' = KE$ where E is the endomorphism of the preceding exercise. Thus since $\mu(g)N' = \{0\}$ we have for all $X \in g$ that

$$0 = \mu(X)E = [\sigma(X), E]$$

and therefore $\ker E$ is $\sigma(g)$-invariant. Thus from exercise (2iii) $V = W + \ker E$ is a $\sigma(g)$-invariant direct sum decomposition so that σ is completely reducible.

Lemma 12.29 Let g be a semisimple Lie algebra, let ρ be a representation of g in a finite-dimensional vector space M, and let N be a subspace of co-dimension 1 so that for every $X \in g$ we have $\rho(X)M \subset N$. Then there exists a direct sum complement of N which is annihilated by $\rho(g)$.

Proof For $X \in g$ let $\sigma(X) = \rho(X) \mid N$. Then we have the following two cases.

Irreducible case Assume σ is simple; that is, N is irreducible. If $\sigma = 0$, then for X, $Y \in g$ we have using the hypothesis

$$\rho(X)\rho(Y)M \subset \rho(X)N = \sigma(X)N = \{0\}$$

so that $\rho(X)\rho(Y) = 0$. Thus since g is semisimple, $g = [gg]$ and therefore $\rho(g) = \rho([gg]) = [\rho(g), \rho(g)] = \{0\}$. Consequently $\sigma = 0$ implies $\rho = 0$.

If $\sigma \neq 0$, let $k = \ker \sigma$. Then k is an ideal of g and since g is semisimple, there exists an ideal h of g so that we have the direct sum of ideals

$$g = h \oplus k.$$

Now $h \neq \{0\}$ otherwise $g = k$ so that $\sigma = 0$ and also $\rho | h$ is injective. To see the latter let $X \in h$ and suppose $\rho(X) = 0$. Then $\sigma(X) = \rho(X) | N = 0$. Thus $X \in \ker \sigma = k$ and therefore $X \in h \cap k = \{0\}$. Using this we have from Corollary 12.22 that the bilinear form

$$\tau(X, Y) = \text{trace } \rho(X)\rho(Y)$$

for $X, Y \in h$ is nondegenerate and from Lemma 12.27 we obtain the Casimir operator $C \in \text{End}(M)$. From the formula

$$C = \sum_i \rho(X_i)\rho(X_i')$$

for $X_i, X_i' \in h$ we have, using the hypothesis $\rho(g)M \subset N$, that

$$C(M) \subset N.$$

Therefore by extending a basis of N to a basis of M we obtain

$$\text{trace}(C | N) = \text{trace } C = \dim h \neq 0$$

so that $C | N \neq 0$. From this and our assumption that N is irreducible we obtain $C | N$ as an automorphism. Thus if we let $P = \ker C$, we obtain

$$P \cap N = \{0\},$$

but from $C(M) \subset N$ we obtain $N = \text{Image}(C)$ and the direct sum

$$M = N + P.$$

However, from Lemma 12.27 we have $[\rho(A), C] = 0$ for all $A \in g$ and therefore P is $\rho(g)$-invariant. This with the hypothesis gives $\rho(g)P \subset P \cap N = \{0\}$; that is, P is a direct complement of N annihilated by $\rho(g)$.

General case We do this case by induction on the dimension of M. In the $\sigma(g)$-module N let S be a nonzero minimal $\sigma(g)$-submodule; that is, $S \subset N$ and S is an irreducible $\sigma(g)$-submodule of M. If $S = N$, then we are in the preceding case; otherwise let

$$M' = M/S \qquad \text{and} \qquad N' = N/S.$$

Now for every $X \in g$ the mapping $\rho(X)$ induces an endomorphism $\rho'(X) : M' \to M'$ and the mapping

$$\rho' : g \to gl(M') : X \to \rho'(X)$$

is a representation of g in the space M'. By the hypothesis we have

$$\rho'(g)(M') \subset N'$$

and by the isomorphism $(M/S)/(N/S) \cong M/N$ we have N' is of codimension 1 in M'. Thus by induction we can assume there exists a subspace $P' \subset M'$ of dimension 1 such that

$$M' = N' + P' \quad \text{(direct sum)} \qquad \text{and} \qquad \rho'(g)P' = \{0\}.$$

Now let $P \subset M$ be the inverse image of P' relative to the homomorphism $M \twoheadrightarrow M'$. Then since $P \supset S$ and

$$1 = \dim P' = \dim P/S \qquad \text{and} \qquad \rho(g)P \subset S$$

we can now apply the first part of this proof to the irreducible module S and its ambient module P. Thus there exists a subspace $Q \subset P$ so that we have the direct sum

$$P = Q + S \qquad \text{and} \qquad \rho(g)Q = \{0\}.$$

From the direct sum $M' = N' + P'$ we have $\{0'\} = N' \cap P'$ so that

$$S = N \cap P \subset N.$$

Also from $M' = N' + P'$ we have for any $A \in M$ that there exist $B \in N$, $C \in P$, and $D \in S$ so that

$$A = B + C + D = C + B'$$

for some $B' \in N$ (using $S \subset N$). However, from the decomposition $P = Q + S$ we can write $C = F + H$ for $F \in Q$ and $H \in S \subset N$. Thus we obtain $A = F + B''$ where $F \in Q$ and $B'' \in N$; that is

$$M = Q + N.$$

This sum is direct: if $Z \in Q \cap N \subset P \cap N = S$, then $Z \in Q$ and $Z \in S$; that is, $Z \in Q \cap S = \{0\}$ using the direct decomposition $P = Q + S$. Thus we have the desired direct sum

$$M = Q + N \qquad \text{with} \quad \rho(g)Q = \{0\}.$$

Theorem 12.30 (Weyl's theorem) Let g be a finite-dimensional Lie algebra over the field K of characteristic 0. Then g is semisimple if and only if every g-module is completely reducible.

PROOF Combining Lemmas 12.28 and 12.29 we find that semisimplicity implies complete reducibility of modules. For the converse we note the hypothesis shows that the adjoint representation is completely reducible. Thus every ideal h of g has a complementary ideal k such that $g = h \oplus k$ and the map

$$\phi : g \to g/k \cong h$$

is a representation of g. Now if g is not semisimple, then from Section 10.2, g has an Abelian ideal h. Writing the direct sum $h = KX_1 + \cdots + KX_n$ for a basis $X_1, \ldots, X_n$ of h we have that the map ψ

$$\psi : h \to KX_1 \cong K$$

is a representation of h (using $[hh] = \{0\}$). Thus

$$\psi \circ \phi : g \to K$$

is a representation of g. However, the one-dimensional algebra K has a representation

$$\rho : K \to gl(2, K) : x \to \begin{bmatrix} 0 & 0 \\ x & 0 \end{bmatrix}$$

which is not completely reducible. Thus $\rho \circ \psi \circ \phi$ is not a completely reducible representation of g. This contradiction shows that no Abelian ideals exist in g; that is, g is semisimple.

Exercises (3) Show the above representation ρ is not completely reducible.

(4) Let h be a semisimple Lie subalgebra of the Lie algebra g. Show there exists a subspace m of g such that $g = m + h$ (direct sum) and $[hm] \subset m$; that is, (g, h) is a reductive pair. Note Section 12.2, exercise (7).

REMARKS (1) If g is a semisimple Lie algebra of endomorphisms acting on the finite-dimensional vector space V over a field K, then V is completely reducible relative to g. Consequently the associative algebra g^* generated by the endomorphism in g is a semisimple associative algebra (Proposition 9.15). Thus results on representations of semisimple associative algebras can be applied to representations of semisimple Lie algebras. For example, let g be a Lie subalgebra of $gl(V)$. Then from complete reducibility we see g is completely reducible in V if and only if g^* is completely reducible in V. However, using exercise (5) of Section 12.1, g^* is completely reducible in V if and only if $P \otimes_K g^*$ is completely reducible in $P \otimes_K V$. Combining these we obtain the following which allows one to assume algebraic closure of the field when discussing complete reducibility.

Proposition 12.31 Let g be a Lie subalgebra of $gl(V)$ where V is a vector space over K and let P be the algebraic closure of K. Then g is completely reducible in V if and only if $P \otimes_K g$ is completely reducible in $P \otimes_K V$.

We now consider the explicit form of a completely reducible Lie algebra of endomorphisms.

Theorem 12.32 Let V be a nonzero finite-dimensional vector space over a field K of characteristic 0 and let g be a Lie subalgebra of $gl(V)$. Then g is completely reducible in V if and only if

(a) $g = c \oplus g'$ (direct sum), where c is the center of g and $g' = [g, g]$ is semisimple or $\{0\}$;
(b) all the endomorphisms of c are semisimple.

We divide the proof into several lemmas.

Lemma 12.33 Let V be a nonzero finite-dimensional vector space over the field K and let $g \subset gl(V)$ act irreducibly in V. Let h be an ideal of g so that every endomorphism of h is nilpotent. Then $h = 0$.

PROOF Using $AX - XA \in h$ for all $A \in g$ and $X \in h$ we see that the subspace hV is a g-submodule of V. By the irreducibility of V we have $hV = \{0\}$ or V. If $hV = V$, then by induction

$$V = hV = h^2V = \cdots = h^kV = \cdots.$$

However by Corollary 11.11 to Engel's theorem we have $h^n = \{0\}$ for suitable n. Thus we obtain the contradiction $V = \{0\}$. Therefore $hV = \{0\}$.

Lemma 12.34 Let V be a finite-dimensional vector space over the field K and let $T \in gl(V)$ be of the form

$$T = \sum_{k=1}^{r} [A_k, B_k]$$

where A_k, $B_k \in gl(V)$ and $[T, A_k] = 0$ for $k = 1, \ldots, r$. Then T is nilpotent.

PROOF From $[T, A_k] = 0$ we have $[T^p, A_k] = 0$ for $p = 0, 1, 2, \ldots$ and therefore for $p = 0, 1, 2\ldots$

$$\begin{aligned} T^{p+1} &= \sum_k T^p(A_k B_k - B_k A_k) \\ &= \sum_k A_k T^p B_k - T^p B_k A_k = \sum_k [A_k, T^p B_k]. \end{aligned}$$

Thus since the trace of any commutator is 0 we have trace $T^{p+1} = 0$ for $p = 0, 1, \ldots.$ By Lemma 12.13 we obtain T as nilpotent.

Lemma 12.35 Let g be a Lie algebra over the field K, let $g' = [gg]$, and let c be the center of g. Assume

(a) $c \cap g' = \{0\}$ and
(b) all Abelian ideals of g are contained in c.

Then $g = c \oplus g'$ (direct sum) where g' is semisimple or $\{0\}$.

PROOF Assume g' is not $\{0\}$ and using (a) let $p = c \oplus g'$ and let q be a complementary subspace of p in g. Thus we have the direct sums of subspaces

$$g = p + q = c + g' + q$$

and we let $m = g' + q$. Therefore $m \cap c = \{0\}$ and since $m \supset g'$ we have $[gm] \subset [gg] \subset m$ so that m is an ideal of g. Therefore $g = c \oplus m$ as a direct sum of ideals. To show that m is semisimple let h be an Abelian ideal of m. Then using $g = c \oplus m$ we have

$$[gh] \subset [ch] + [mh] \subset h$$

so that h is an Abelian ideal of g. However, from assumption (b) we have $h \subset c$ and therefore $h \subset c \cap m = \{0\}$. Consequently by Corollary 12.23(b), m is semisimple. Thus since $m = [mm]$ we have $g' = [gg] = [c + m\ c + m] = [mm] = m$.

PROOF OF THEOREM 12.32. We first show conditions (a) and (b) above hold for the irreducible case, for let g act irreducibly on V and let

$$T \in c \cap g'.$$

Then $T = \sum [A_k, B_k]$ and $[T, A_k] = 0$ (since $T \in c$). By Lemma 12.34 we have T as nilpotent. Thus since every element of the ideal $c \cap g'$ is nilpotent we have by Lemma 12.33 that $c \cap g' = \{0\}$. Thus condition (a) of Lemma 12.35 is satisfied. For condition (b) let h be an Abelian ideal of g and let $T \in [h, g]$ which is an ideal of g. Then

$$T = \sum [A_k, B_k]$$

for $A_k \in h$ and $B_k \in g$. Since h is an ideal we have $T \in h$ and since h is Abelian, $[T, A_k] = 0$. Thus, by Lemma 12.34, T is nilpotent and as before, from Lemma 12.33, $[h, g] = \{0\}$. Therefore $h \subset c$.

Next assume V is g-completely reducible with direct sum decomposition

$$V = V_1 + \cdots + V_r$$

where each V_i is g-irreducible. The set

$$k_i = \{A \in g : AV_i = \{0\}\}$$

is an ideal in g and note that $\bigcap_{i=1}^r k_i = \{0\}$. Let $\bar{g}_i = g/k_i$ and let $\bar{c}_i$ be the center of $\bar{g}_i$. Then V_i is an irreducible $\bar{g}_i$-module (where the action is defined by $\overline{T}X = TX$ for $\overline{T} = T + k_i \in \bar{g}_i$ and $X \in V_i$). Now conditions (a) and (b) hold for $\bar{g}_i$ on V_i; that is, for the irreducible case, and they hold for g on V as follows.

(a) Let $A \in g' \cap c$. Then for each $i = 1, \ldots, r$

$$A + k_i \in \bar{g}_i' \cap \bar{c}_i = \{\bar{0}_i\}.$$

Thus $A \in k_i$, that is, $A \in \bigcap k_i = \{0\}$ so that $g' \cap c = \{0\}$.

(b) If h is an Abelian ideal of g, then $\bar{h} = h + k_i$ is an Abelian ideal of $\bar{g}_i$. Therefore $\bar{h} \subset \bar{c}_i$ so that $[\bar{h}, \bar{g}_i] = \{\bar{0}_i\}$. Thus $[h, g] \subset k_i$ for $i = 1, \ldots, r$ which yields $[h, g] \subset \bigcap k_i = \{0\}$ and therefore $h \subset c$.

We now show that the endomorphisms of c are semisimple, for suppose $A \in c$ is not semisimple. Then from Section 12.2 the algebra $K[A]$ generated by A has a nilpotent ideal R. However, since A commutes with every endomorphism of the associative algebra g^*, the set $N = Rg^*$ is an ideal of g^* which is nilpotent. To see the latter we just note $N^2 = Rg^* \cdot Rg^* \subset R^2 g^*$ so by induction $N^k \subset R^k g^* = \{0\}$ for large enough k. Since V is a completely reducible g-module, V is also a completely reducible g^*-module. Thus from Proposition 9.15, g^* is semisimple and therefore we must have $N = \{0\}$; that is, $A = 0$.

To show the converse part of Theorem 10.32 we assume conditions (a) and (b) hold and note from Proposition 12.31 we can assume the field K is algebraically closed. By (b) and Theorem 11.14 we have the weight space direct sum decomposition

$$V = V(\phi_1) + \cdots + V(\phi_n)$$

where $V(\phi_k) = \{X \in V : \text{for all } T \in c \text{ implies } (T - \phi_k(T)I)X = 0\}$. By (a) we have c is the center of $g = c \oplus g'$ and therefore the weight spaces $V(\phi_k)$ are g-invariant. However, the g-submodules of each $V(\phi_k)$ are the same as the g'-submodules. Thus if g' is semisimple, we can apply Weyl's theorem (12.30) to conclude each $V(\phi_k)$ is g-completely reducible. By the above direct sum decomposition V is g-completely reducible.

If g' is $\{0\}$, then each of the weight spaces $V(\phi_k)$ is c-completely reducible (why?) so that V is actually g-completely reducible.

REMARK (2) The preceding direct sum decomposition $g = c \oplus g'$ has the following generalization; see the proofs of Jacobson [1962] and Serre [1965].

Theorem (Levi–Malcev) Let g be a finite-dimensional Lie algebra over the field K and let $r = \operatorname{rad} g$.

(a) There exists a subalgebra h of g such that

$$g = r + h$$

is a subspace direct sum. Furthermore $h \cong g/r$ is semisimple or $\{0\}$. The decomposition $g = r + h$ is called a **Levi decomposition**.

(b) Let $g = r + h = r + k$ be two Levi decompositions of g. Then there exists $\phi \in \operatorname{Aut}(g)$ such that $\phi(h) = k$. Furthermore the automorphism ϕ can be chosen to be of the form $\phi = \exp(\operatorname{ad} X)$ where $X \in [gg] \cap r$.

We can use this to prove the following variation of a result of Lie [Cohn, 1957, p. 99; Hochschild, 1965, p. 133].

Theorem Let g be a finite-dimensional Lie algebra over R. Then there exists a real Lie group G whose Lie algebra is g.

Briefly, we write a Levi decomposition $g = r + h$ and note that this is a semidirect sum. Next we use the result of Section 10.2 which states that the solvable Lie algebra r is the Lie algebra of a solvable Lie group R. Thus since the semisimple algebra h is such that the adjoint representation is faithful, then the group H generated by $\exp(\operatorname{ad} h)$ has a Lie algebra isomorphic to h. Therefore the semidirect product $G = R \times H$ has Lie algebra (isomorphic to) g.

The above result also follows from Ado's theorem which states: If g is a finite-dimensional Lie algebra over the field K, then g has a faithful finite-dimensional representation. Thus if $\rho : g \to gl(V)$ is such a faithful representation of a real Lie algebra, then the subgroup G generated by $\exp \rho(g) \subset GL(V)$ has Lie algebra (isomorphic to) g.

5. More on Radicals, Derivations, and Tensor Products

In this section we use preceding results to discuss relationships between radicals of Lie and associative algebras, the complete reducibility of tensor products, and derivations of simple nonassociative algebras. Following Jacobson [1962] we have the next lemma; compare with Lemma 12.33.

Lemma 12.36 Let V be a finite-dimensional vector space over the field K, let $g \subset gl(V)$ be a Lie algebra of endomorphisms, and let h be an ideal of g such that every endomorphism in h is nilpotent. Then the associative algebra h^* generated by h is contained in the radical N of the associative algebra g^*.

PROOF The subspace $n = g^*h^* + h^*$ is an ideal of g^* noting first

$$g^*n \subset g^*(g^*h^*) + g^*h^* \subset g^*h^* \subset n.$$

Next using $AX - XA \in h$ for all $A \in h$, $X \in g$ we have by an easy induction

$$\underbrace{hh\ldots h}_{k}\, g \subset g\, \underbrace{h\ldots h}_{k} + \underbrace{h\ldots h}_{k}.$$

Therefore $h^*g^* \subset g^*h^* + h^*$ and

$$\begin{aligned} ng^* &\subset g^*h^*g^* + h^*g^* \\ &\subset g^*(g^*h^* + h^*) + g^*h^* + h^* \subset n. \end{aligned}$$

Also $n^k \subset g^*h^* + (h^*)^k$ so that from Corollary 11.11 to Engel's theorem h^* is nilpotent and therefore for k large enough $n^k \subset g^*h^*$. Now by induction we can show

$$(h^*)^r g^* \subset g^*(h^*)^r + (h^*)^r$$

and consequently

$$(g^*h^*)^r \subset g^*(h^*)^r.$$

Using this and h^* as nilpotent we see that $(n^k)^r \subset g^*(h^*)^r = \{0\}$ for suitably large r. Thus the nilpotent ideal n is contained in N; that is, $h^* \subset n \subset N$.

Lemma 12.37 Let V be a finite-dimensional vector space over the field K and let $g \subset gl(V)$ be a solvable Lie algebra. If N is the radical of g^*, then g^*/N is a commutative (associative) algebra.

PROOF First assume g^* is semisimple. Then V is a completely reducible g^*-module (Theorem 9.16) so that V is a completely reducible g-module. Thus from Theorem 12.32 we conclude that $g = c \oplus g'$ where g' is $\{0\}$ or semisimple (and therefore $g' = [g', g']$). However, since g is solvable we must have $g' = \{0\}$. Thus $g^*/N = g^*/\{0\} = c^*$ is a commutative algebra. In particular, g is an Abelian Lie algebra.

For the general case we note that since N is an ideal of g^* and since as sets $g^* \subset gl(V) = \mathrm{End}(V)$ we can regard N as solvable Lie subalgebra of $gl(V)$; still denoted by N. Thus since g is solvable, the Lie algebra $(g + N)/N$ is solvable and the associative algebra generated by this Lie algebra is g^*/N which is semisimple. Thus from the preceding paragraph $(g + N)/N$ is an Abelian Lie algebra which means in terms of cosets $(X + N)(Y + N) = (Y + N)(X + N)$ for endomorphisms $X, Y \in g$. However, these cosets generate g^*/N so that g^*/N is commutative.

Theorem 12.38 Let g be a Lie algebra of endomorphisms acting on the finite-dimensional vector space V over the field K and let r be the radical of g and N be the radical of g^*. Then:

(a) $g \cap N$ is the set of nilpotent endomorphisms in r;
(b) $[r, g] \subset N$.

PROOF Let r_0 be the set of nilpotent endomorphisms of r. Then since N is an associative nilpotent ideal, we have $g \cap N \subset r$. Since the endomorphisms of N are nilpotent $g \cap N \subset r_0$. If rad(r^*) denotes the radical of the associative algebra r^* generated by r, then from Lemma 12.37, $r^*/\text{rad}(r^*)$ is commutative. However, this implies $r^*/\text{rad}(r^*)$ has no nonzero nilpotent elements as follows. A nonzero nilpotent element in the commutative algebra generates a nilpotent ideal and therefore is in the radical of the algebra; but $r^*/\text{rad}(r^*)$ has 0 radical. Thus rad(r^*) equals the set of nilpotent elements of r^* and therefore $r_0 \subset \text{rad}(r^*)$. From this we obtain

$$r_0 = r \cap \text{rad}(r^*)$$

and in particular r_0 is a subspace of g.

We shall next show r_0 is an ideal of the Lie algebra g. For this we use the characterization of the radical of an associative algebra A as

$$\text{rad } A = \{Z \in A : f(Z, X) = 0 \text{ for all } X \in A\} \tag{$*$}$$

where $f(X, Y) = \text{trace } L(X)L(Y)$; see Proposition 12.14. First for any $X, Y, Z \in A$ we have

$$f([Z, X], Y) = f(Z, [X, Y])$$

where $[Z, X] = ZX - XZ$. Now let $A = r^*$ and since r is an ideal in g we have for $Z \in g$ and $X, Y \in r$ that

$$[Z, XY] = [Z, X]Y + X[Z, Y] \in r^*$$

so that by induction

$$[g, r^*] \subset r^*.$$

Now let $U \in r_0 = r \cap \text{rad}(r^*)$ and let $Z \in g$. Then $[Z, U] \in r$ and we now show $[Z, U] \in \text{rad}(r^*)$ as follows. For any $X \in r^*$

$$f(X, [Z, U]) = f([X, Z], U).$$

However, $[X, Z] \in [r^*, g] \subset r^*$ and $U \in \text{rad}(r^*)$ so that from formula ($*$) we have $f([X, Z], U) = 0$. Thus $f(X, [Z, U]) = 0$ so that from ($*$) we obtain $[Z, U] \in \text{rad}(r^*)$. Therefore $[g, r_0] \subset r_0$.

Since r_0 is an ideal in g all of whose elements are nilpotent we can apply Lemma 12.36 to conclude $r_0 \subset \text{rad}(g^*) = N$. Thus since $r_0 \subset g$ we have $r_0 \subset N \cap g$ and therefore $r_0 = N \cap g$ which proves part (a).

For (b) we use the fact that if L is a Lie algebra of endomorphisms so that L^* is semisimple, then rad L is contained in the center of L; this follows from Theorem 12.32. We apply this to $L = (g + N)/N$ as discussed in Lemma 12.37 and note that the associative algebra generated by L is $L^* = g^*/N$ which is semisimple or $\{\bar{0}\}$. If L^* is $\{\bar{0}\}$, then we are done. Otherwise the radical of $(g + N)/N$ is in the center of $(g + N)/N$. However, $(r + N)/N$ is a solvable ideal of $(g + N)/N$ and therefore in the radical of $(g + N)/N$. Thus

$$[(r + N)/N, (g + N)/N] = \{\bar{0}\};$$

that is, $[r, g] \subset N$.

Exercise (1) Show how $(g + N)/N$ in the above proof is actually a Lie algebra of endomorphisms.

Corollary 12.39 Let g be a finite-dimensional Lie algebra over the field K. Then

$$\operatorname{rad}(g) = \{X \in g : \operatorname{Kill}(X, [AB]) = 0 \text{ for all } A, B \in g\};$$

that is, $\operatorname{rad}(g) = g'^{\perp}$ relative to the Killing form of g.

PROOF Let $r = \operatorname{rad}(g)$ and let $A \in r$ and $X, Y \in g$. Then $\operatorname{Kill}(A, [XY]) = \operatorname{Kill}([AX], Y)$. However, from Theorem 12.38 we have that

$$\operatorname{ad}[AX] = [\operatorname{ad} A, \operatorname{ad} X] \in [\operatorname{rad}(\operatorname{ad} g), \operatorname{ad} g]$$

which is contained in the radical of the associative algebra $(\operatorname{ad} g)^*$. Thus $\operatorname{ad}[AX] \operatorname{ad} Y$ is in the radical of $(\operatorname{ad} g)^*$ because the radical is an ideal. Since this radical consists of nilpotent endomorphisms,

$$0 = \operatorname{trace} \operatorname{ad}[AX] \operatorname{ad} Y = \operatorname{Kill}(A, [XY]).$$

Thus $A \in g'^{\perp}$ so that $r \subset g'^{\perp}$.

For the other inclusion we first note that $h = g'^{\perp}$ is an ideal of g. Next if $X \in h' = [hh]$, then we have $0 = \operatorname{Kill}(X, X) = \operatorname{trace}(\operatorname{ad} X)^2$. Therefore by Cartan's criterion (Corollary 12.21) we have h is a solvable ideal of g; that is, $g'^{\perp} \subset r$.

Recall [Section 11.2, exercise (5)] that the nilpotent radical n of a finite-dimensional Lie algebra g is the maximal nilpotent ideal of g.

Corollary 12.40 Let g be a finite-dimensional Lie algebra over the field K and let r be its radical and n its nilpotent radical. Then $[rg] \subset n$.

PROOF From Theorem 12.38 we have $[\operatorname{ad} r, \operatorname{ad} g]$ is in the radical of $(\operatorname{ad} g)^*$. Since this radical is associative nilpotent, there exists k so that for any $A_1, \ldots, A_k \in r$ and $X_1, \ldots, X_k \in g$ that

$$\begin{aligned} 0 &= [\operatorname{ad} A_1, \operatorname{ad} X_1] \cdots [\operatorname{ad} A_k, \operatorname{ad} X_k] \\ &= \operatorname{ad}[A_1 X_1] \cdots \operatorname{ad}[A_k X_k]. \end{aligned}$$

Thus by Corollary 11.12, the Lie algebra $[rg]$ is nilpotent. However, $[rg]$ is an ideal of g (using the Jacobi identity) so that $[rg] \subset n$.

We generalize this result as follows:

Corollary 12.41 Let g be a finite-dimensional Lie algebra over the field K and let r be the radical of g and n the nilpotent radical of g. Then every derivation of g maps r into n.

PROOF Let D be a derivation of g. Then from $r = g'^{\perp}$ we have for $X \in r$ and $A, B \in g$ that

$$\mathrm{Kill}(DX, [AB]) = -\mathrm{Kill}(X, D[AB]) = 0$$

where we use

$$D[AB] = [DA\ B] + [A\ DB] \in g' \text{ and } \mathrm{Kill}(DX, Y) = -\mathrm{Kill}(X, DY)$$

from Corollary 12.3. Thus $Dr \subset r$.

Next we shall show that the subspace $h = r' + Dr$ is a nilpotent ideal of r. Clearly h is an ideal of r and let

$$L = r \times KD = r \oplus KD$$

be the (external) direct sum of r and the one-dimensional space KD. Define a multiplication on L which extends the multiplication of r by

$$[X + aD, Y + bD] = [XY] + aDY - bDX.$$

This makes L into a Lie algebra in which r is an ideal. However, L/r is one-dimensional and therefore solvable. Since r is solvable, this implies L is solvable. By Proposition 11.9 $L' = [LL]$ is a nilpotent ideal in L, but

$$L' = r' + Dr$$

is contained in r; that is, the subspace $h = r' + Dr$ is a nilpotent ideal of r.

Next since the nilpotent radical n of g is a nilpotent ideal of r we have that

$$k = h + n$$

is a nilpotent ideal of r. However, k is an ideal of g as follows. Since n and r' are ideals of g it suffices to show $[Dr\ g] \subset k$. For $X \in g$ and $A \in r$ we have

$$[DA\ X] = D[AX] - [A\ DX]$$

where $D[AX] \in D[rg] \subset Dr$ and $[ADX] \in [rg] \subset n$ by Corollary 12.40. Thus $[Dr\ g] \subset Dr + n \subset k$ so that k is actually a nilpotent ideal of g; that is, $r' + Dr + n \subset n$. Since $r' \subset n$ we obtain $Dr \subset n$.

As another application of Theorem 12.32 on complete reducibility we have the following results on derivations due to Jacobson for algebras with an identity element [Schafer, 1966]. For a nonassociative algebra A, recall that the Lie transformation algebra $L(A)$ is the Lie algebra generated by the left and right multiplication endomorphisms of A (Section 7.2).

Proposition 12.42 Let A be a finite-dimensional nonassociative algebra over an algebraically closed field K of characteristic 0 such that A is a direct sum of ideals which are simple subalgebras. If D is a derivation of A, then $D = U + aI$ where $U \in [L(A), L(A)]$, I is the identity endomorphism, and $a \in K$.

PROOF If $A = A(1) \oplus \cdots \oplus A(n)$ is the given direct sum decomposition into simple ideals, then analogous to the proof of Corollary 12.26 we have for $i = 1, \ldots, n$ that

$$DA(i) = D\big(A(i)A(i)\big) \subset A(i)\big(DA(i)\big) \subset A(i).$$

Thus $D : A(i) \to A(i)$ induces a derivation of $A(i)$. Also using the corresponding decomposition $L(A) = L(A(1)) + \cdots + L(A(n))$, it suffices to consider the case when A is a simple algebra.

Let $\mathscr{L} = L(A)$ be the Lie transformation algebra of A which is generated by the left and right multiplications $L(X)$ and $R(Y)$ for all $X, Y \in A$. Since A is simple $\mathscr{L}$ acts irreducibly on A. Therefore according to Theorem 12.32, $\mathscr{L} = C \oplus \mathscr{L}'$. Now since C is the center of $\mathscr{L}$, any endomorphism $T \in C$ commutes with an irreducible set of endomorphisms and therefore $T = aI$ where $a \in K$, using Schur's lemma [Proposition 9.14(a)]. Thus C equals KI or $\{0\}$.

From $[D, L(X)] = L(DX)$ and $[D, R(X)] = R(DX)$ we see $[D, \mathscr{L}] \subset \mathscr{L}$. However, since $\mathscr{L} = C \oplus \mathscr{L}'$ with C as above we obtain $[D, \mathscr{L}'] \subset \mathscr{L}'$. First we assume $\mathscr{L}' \neq \{0\}$. Then the mapping

$$\mathfrak{D} : \mathscr{L}' \to \mathscr{L}' : P \to [D, P]$$

is a derivation of the semisimple Lie algebra $\mathscr{L}'$. Thus there exists $U \in \mathscr{L}'$ such that $\mathfrak{D} = \operatorname{ad} U$. Thus for any $V = bI + V' \in \mathscr{L} = C \oplus \mathscr{L}'$ we see

$$[D, V] = [D, V'] = \mathfrak{D}V' = \operatorname{ad} U(V') = [U, V]$$

so that $[D - U, V] = 0$. Thus $D - U$ commutes with an irreducible set of endomorphisms which implies $D - U = aI$.

In case $\mathscr{L}' = \{0\}$ we see $\mathscr{L} = KI$. Thus for any $X \in A$, $L(X) = \alpha I$ and $R(X) = \beta I$ so that $0 = [D, L(X)] = L(DX)$ and similarly $R(DX) = 0$. Since A is simple, this implies $DX = 0$; that is, $D = 0$ which is in $\mathscr{L}$.

Exercises (2) Let A be a simple nonassociative algebra over K and let $X \in A$ be such that $L(X) = R(X) = 0$. Show $X = 0$.

(3) Let A be a simple nonassociative algebra over an algebraically closed field of characteristic 0 and let D be a derivation of A.

(i) If A contains an identity element, then show that $D \in L(A)$.

(ii) If trace $D = 0$, then show that $D \in L(A)$. [This is the case for Lie algebras since $\operatorname{Kill}(DX, Y) = -\operatorname{Kill}(X, DY)$.]

Recall from Section 9.4 that if g is a Lie algebra over the field K and V and W are g-modules with corresponding representations ρ and σ, then the vector space tensor product $V \otimes W$ becomes a g-module where the action of g is defined by

$$A(\sum X \otimes Y) = \sum \rho(A)X \otimes Y + X \otimes \sigma(A)Y.$$

Proposition 12.43 Let g be a Lie algebra over the field K and let V and W be completely reducible g-modules. Then the tensor product $V \otimes W$ is a completely reducible g-module.

We outline a proof as follows. From Proposition 12.31 we can assume K is algebraically closed.

Exercises (4) Show the (external) direct sum $P = V \oplus W$ is a g-module which is completely reducible. Also show $V \otimes W$ is a submodule of $P \otimes P$. Thus since a submodule of a completely reducible module is also completely reducible, it suffices to show $P \otimes P$ is completely reducible.

(5) (i) Let τ be the completely reducible representation of g in P and assume τ is faithful. Show that $g = c \oplus g'$ (direct sum) where c is the center of g and $g' = [gg]$ is semisimple or $\{0\}$.

(ii) Show that it suffices to assume τ is faithful in (i) for the proof of the proposition.

(6) (i) Let $\tilde{\tau}$ be the representation of g in $P \otimes P$ and show that $\tilde{\tau}(g) = \tilde{\tau}(c) \oplus \tilde{\tau}(g')$ where $\tilde{\tau}(c)$ is the center of $\tilde{\tau}(g)$ and $\tilde{\tau}(g')$ is semisimple or $\{0\}$.

(ii) Recalling K is algebraically closed, show that for any $A \in c$, the endomorphism $\tilde{\tau}(A) = \tau(A) \otimes I + I \otimes \tau(A)$ is semisimple. [If X, $Y \in K$ are characteristic roots of $\tau(A)$, what are the characteristic roots of $\tilde{\tau}(A)$?] Thus by Theorem 12.32 conclude the Proof of Proposition 12.43.

6. Remarks on Real Simple Lie Algebras and Compactness

We combine the results of Section 9.2 on complexification with some of the results on semisimplicity to obtain the form of a real simple Lie algebra. Analogous to the last part of the proof of Cartan's criterion for solvability, we have the following result.

Lemma 12.44 Let g be a Lie algebra over K, and let P be the algebraic closure of K. Then g is semisimple over K if and only if $\tilde{g} = P \otimes_K g$ is a semisimple Lie algebra over P.

PROOF Let $X_1, \ldots, X_m$ be a basis of g over K, and let $b_{ij} = \text{Kill}(X_i, X_j)$, let $\bar{X}_1, \ldots, \bar{X}_m$ be the corresponding basis for $\tilde{g}$, and let $\bar{b}_{ij} = \text{Kill}(\bar{X}_i, \bar{X}_j)$. Then using the computations analogous to the last part of the proof of Cartan's criterion for solvability [preceding exercise (2)], we see $b_{ij} = \bar{b}_{ij}$. Thus g (respectively $\tilde{g}$) is semisimple if and only if its Killing form is nondegenerate which is the case if and only if the matrix (b_{ij}) [respectively $(\bar{b}_{ij})$] is nondegenerate. This yields the result, since $b_{ij} = \bar{b}_{ij}$.

For a real Lie algebra g, we see that g is simple over R implies $\tilde{g} = C \otimes g$ is semisimple over C. From Section 9.2 we notice that $\tilde{g}$ need not be simple. However, we have the following result.

Theorem 12.45 Let g be a simple Lie algebra over the reals R. Then g is isomorphic to the realification of a simple complex Lie algebra, or g is isomorphic to a real form of a simple complex Lie algebra.

PROOF If $\tilde{g} = C \otimes g$ is simple, then g is a real form for the simple complex Lie algebra $\tilde{g}$. For the other case we have $\tilde{g}$ is not simple, but, from Lemma 12.44 $\tilde{g}$ is semisimple. Let h be a proper simple ideal in $\tilde{g}$, and let C be the conjugation given by $C : \tilde{g} \to \tilde{g} : X + iY \to X - iY$. Then it is easy to see that $h \cap C(h)$ is an ideal in g, and since g is simple, $h \cap C(h) = \{0\}$ or $h \cap C(h) = g$. In the first case, the set $\{X + C(X) : X \in h\}$ is an ideal in g and therefore equals g. Thus $\tilde{g} = h \oplus C(h)$ using $h \cap C(h) = \{0\}$. Therefore $g \oplus g \cong \tilde{g}_R \cong h_R \oplus (C(h))_R \cong h_R \oplus h_R$. This shows g is isomorphic to the realification h_R of the simple complex algebra h. The case $h \cap C(h) = g$ is left as an exercise.

REMARK (1) From Theorem 12.45 we see that in order to find the real semisimple Lie algebras, it suffices to find the simple complex Lie algebras and then find their realifications and real forms. We do this in Chapter 15 noting that the "realification" part of the problem can be done as follows.

Proposition 12.46 Let g be a simple real Lie algebra which is not isomorphic to a real form of a simple complex Lie algebra. Then g is isomorphic to the realification of a simple complex Lie algebra, and conversely every realification of a simple complex Lie algebra is a simple real Lie algebra.

PROOF It suffices to show the converse. Thus let $\tilde{g}$ be a simple complex Lie algebra, then clearly $\tilde{g}_R$ is semisimple. Now let h be a nonzero ideal in $\tilde{g}_R$. Then h is semisimple, and therefore $h = [hh]$. Now for $X, Y \in h \subset \tilde{g}$ we have by definition of scalar multiplication in $\tilde{g}$ that $i[XY] = [iXY] \in [\tilde{g}h] \subset h$ noting $\tilde{g} = \tilde{g}_R$ as sets. Thus $ih = i[hh] \subset h$ so that h is closed relative to scalar multiplication by complex numbers. Thus h is an ideal of the simple complex Lie algebra $\tilde{g}$ so that $h = \tilde{g}$; that is, $h = \tilde{g}_R$.

REMARK (2) We shall see in Section 15.1 how real forms are related to compactness, which we now discuss.

Theorem Let G be a connected real semisimple Lie group with Lie algebra $\mathfrak{g}$.

(a) If G is compact, then its universal covering group is compact.

(b) The group G is compact if and only if the Killing form of $\mathfrak{g}$ is negative definite.

For the proofs we refer to the work of Hausner and Schwartz [1968] and Helgason [1962]. Thus we say a Lie algebra is **compact** if its Killing form is negative definite. For example, the compact Lie group $SO(n)$ has Lie algebra $so(n)$ of skew-symmetric matrices and $\text{Kill}(X, Y) = (n-2)$ trace XY. However, since $0 \neq X$ is skew-symmetric, we see trace $XX < 0$.

CHAPTER 13

CARTAN SUBALGEBRAS AND ROOT SPACES

In this chapter we start the detailed analysis of semisimple Lie algebras over a field K of characteristic 0. In the first section the Cartan subalgebra h of a Lie algebra g is introduced. Since h is a nilpotent Lie subalgebra of g, we consider the weight space decomposition of g relative to the nilpotent Lie algebra of endomorphisms $\mathrm{ad}_g\, h$. Thus in the second section we consider the relationships between these weight spaces for a split semisimple Lie algebra and give examples. It turns out that a split semisimple Lie algebra g is a sum of simple subalgebras which are isomorphic to $sl(2, K)$. Consequently, in the third section we discuss the representations of $sl(2, K)$ and apply this to a further analysis of the weight spaces of g.

1. Cartan Subalgebras

All Lie algebras in this chapter will be assumed to be finite-dimensional over a field K of characteristic 0. Recall that a Lie algebra g over K is said to be "split" if for each $X \in g$ all the characteristic values of ad X are in K. All algebras over an algebraically closed field, in particular the complex numbers, are split.

Definition 13.1 Let g be a Lie algebra over K.

(a) For any subalgebra h of g let $N(h) = \{X \in g : [Xh] \subseteq h\}$, then $N(h)$ is called the **normalizer** of h and is the largest subalgebra of g which contains h as an ideal.

(b) A subalgebra h of g is called a **Cartan subalgebra** if it is nilpotent and $N(h) = h$.

For any $X \in g$ and $\lambda \in K$ set

$$g(\lambda, X) = \{Y \in g : (\text{ad } X - \lambda I)^n Y = 0 \text{ for some } n\}.$$

Notice that if λ is a weight of ad X, then $g(\lambda, X)$ is merely a weight space of g for the linear transformation ad X. If $\lambda_1, \ldots, \lambda_m$, the weights of ad X, all lie in K, then it is clear that $g = g(\lambda_1, X) + \cdots + g(\lambda_m, X)$ as a vector space direct sum; recall Section 10.3. From exercise (2) of Section 11.3 we have that if $\lambda + \mu$ is a weight, then

$$[g(\lambda, X)g(\mu, X)] \subseteq g(\lambda + \mu, X)$$

which implies $g(0, X)$ is a subalgebra of g and $[g(0, X)g(\mu, X)] \subset g(\mu, X)$. Briefly, to see this we note that $D = \text{ad } X$ is a derivation of g. Then for $U \in g(\lambda, X)$ and $V \in g(\mu, X)$

$$(D - (\lambda + \mu)I)[UV] = [(D - \lambda I)U\, V] + [U\,(D - \mu I)V].$$

Thus by induction we obtain

$$(D - (\lambda + \mu)I)^p[UV] = \sum_{k=0}^{p} \binom{p}{k}[(D - \lambda I)^k U (D - \mu I)^{p-k} V]$$

and for p large enough, the right side of this equation is 0; that is, $[UV] \in g(\lambda + \mu, X)$. Note that if $\lambda + \mu$ is not a weight of ad X, then

$$[g(\lambda, X)g(\mu, X)] = \{0\}.$$

An element $X \in g$ is called a **regular** element of g if the dimension of $g(0, X)$ is minimal. This minimal dimension is called the **rank** of g.

Example (1) Let $g = gl(2, K)$ the Lie algebra of 2×2 matrices over K. As usual let I denote the identity matrix

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

One can compute that $g(0, I) = g(0, E) = g$ so I and E are not regular. But $g(0, H) = KI + KH$ which is equal to the diagonal matrices of g and it is not too difficult to show that H is regular, $g(0, H)$ is a Cartan subalgebra of g, and so $gl(2, K)$ has rank 2.

Proposition 13.2 Let g be any finite-dimensional Lie algebra over a field K of characteristic 0, $X \in g$ a regular element, and $h = g(0, X)$. Then h is a Cartan subalgebra of g.

PROOF We must show that h is nilpotent and that $N(h) = h$. By Engel's theorem we can show that h is nilpotent by showing that for all $H \in h$, ad H acts as a nilpotent linear transformation on h.

Let $p(t) = \det(\text{ad } X - tI) = t^r q(t)$ be the characteristic polynomial of ad X where $r = \dim h$ is the rank of g and t and does not divide $q(t)$. We then have $h = \{Y \in g : (\text{ad } X)^r Y = 0\}$ and if we let $k = \{Y \in g : q(\text{ad } X)(Y) = 0\}$, the primary decomposition theorem of linear algebra tells us that $g = h + k$, a vector space direct sum, and that h and k are invariant under ad X; that is, $(\text{ad } X)h \subset h$ and $(\text{ad } X)k \subset k$. In order to show that h is nilpotent and $N(h) = h$ we can extend the field K to its algebraic closure if need be and so we may assume that all the weights of ad X lie in K. In this case $k = \sum_{\lambda \neq 0} g(\lambda, X)$ where the sum is over nonzero weights of ad X.

Now assume $Y = H + Y_1 + Y_2 + \cdots + Y_m \in N(h)$ where $Y_i \in g(\lambda_i, X)$, $\lambda_i \neq 0$, and $H \in h = g(0, X)$. Since $X \in h$ we have

$$[XY] = [XH] + [XY_1] + \cdots + [XY_m] \in h.$$

Thus using the above direct sum $g = h + k$ and using the ad X-invariance of $g(\lambda_i, X)$ we see $[XY_i] = 0$. Thus $Y_i \in g(\lambda_i, X) \cap h = \{0\}$. Consequently $Y = H \in h$ and $N(h) = h$.

To show that ad H is nilpotent on h for all $H \in h$ we first choose a basis $X_1, X_2, \ldots, X_n$ for g with $X_1, X_2, \ldots, X_r \in h$ and $X_{r+1}, \ldots, X_n \in k$. With respect to this basis the matrices associated with the linear transformations ad X and ad H look like

$$\text{ad } X = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \qquad \text{ad } H = \begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix}$$

where we know that A is an $r \times r$ nilpotent matrix and B is an $(n - r) \times (n - r)$ nonsingular matrix, and we want to show that the $r \times r$ matrix C is nilpotent. Assume to the contrary that C is not nilpotent and let x, y, and t be three indeterminants. Define

$$p(x, y, t) = \det(x \,\text{ad}\, X + y \,\text{ad}\, H - tI) = \det(xA + yC - tI)\det(xB + yD - tI).$$

Then B nonsingular implies t does not divide $\det(B - tI)$ nor does t divide $\det(xB + yD - tI)$, and C not nilpotent implies t^r does not divide $\det(C - tI)$ nor does t^r divide $\det(xA + yC - tI)$. Thus t^r does not divide $p(x, y, t)$.

However, then one can find $\alpha, \beta \in K$ so that t^r does not divide $p(\alpha, \beta, t) = \det(\mathrm{ad}(\alpha X + \beta H) - tI)$ which is the characteristic polynomial of $\mathrm{ad}(\alpha X + \beta H)$. This contradicts the fact that X is regular and completes the proof.

Corollary 13.3 Every Lie algebra possesses Cartan subalgebras.

Proposition 13.4 If h is a Cartan subalgebra of a Lie algebra g and $X \in h$ with X regular, then $h = g(0, X)$.

PROOF For any $Y \in h$, $(\mathrm{ad}\ X)^m Y = 0$ for some m because h is nilpotent so $Y \in g(0, X)$ and $h \subseteq g(0, X)$.

Now assume that $h \neq g(0, X)$. Then $N(h) = h$ implies that $(\mathrm{ad}\ Z)h \not\subseteq h$ for any $Z \in g(0, X)$ with $Z \notin h$. However, then by induction, for any positive integer n, there exist $H_1, H_2, \ldots, H_n \in h$ with

$$[\cdots[[ZH_1]H_2]\cdots H_n] \notin h$$

and this contradicts the nilpotence of $g(0, X)$; note Proposition 13.2.

Proposition 13.5 If g is a split Lie algebra over K and h is a Cartan subalgebra of g, then $h = g(0, X)$ for some regular $X \in h$.

PROOF By the previous proposition we need only show that h contains a regular element. Since g is split and since $\mathrm{ad}_g(h)$ is nilpotent $g = h + g_1 + \cdots + g_m$, a vector space direct sum of weight spaces of the weight functions of $\mathrm{ad}_g(h)$; Section 11.3. The condition $N(h) = h$ guarantees that h is precisely the weight space of the weight function which is identically 0 on $\mathrm{ad}_g(h)$. Thus for each $i = 1, 2, \ldots, m$ there exists an $H_i \in h$ with ad H_i acting as a nonsingular linear transformation on g_i. Now a standard argument over infinite fields can be used to choose an $H \in h$ equal to a linear combination of the H_i's such that ad H is nonsingular on $g_1 + \cdots + g_m$. Then $g(0, H) = h$ and H is regular as required.

REMARKS (1) This proposition can be used to help prove that if K is algebraically closed, then any two Cartan subalgebras of g are conjugate by an automorphism of g [Jacobson, 1962, p. 273; Hausner and Schwartz, 1968].

(2) In the next section we consider split semisimple Lie algebras so by Proposition 13.5 in this case all Cartan subalgebras will contain regular elements and the dimension of any Cartan subalgebra will be equal to the rank of the algebra.

(3) Let h be a Cartan subalgebra of g. Then since h is nilpotent, $\mathrm{ad}_g\, h$ is a nilpotent Lie algebra of endomorphisms acting on g, for if n is an integer such that for all $H_1, H_2, \ldots, H_n$ in h we have $[\cdots[H_1H_2]\cdots H_n] = 0$, then

$[\cdots[\mathrm{ad}_g H_1, \mathrm{ad}_g H_2]\cdots, \mathrm{ad}_g H_n] = \mathrm{ad}_g([\cdots[H_1H_2]\cdots H_n]) = 0$. In the next section we use results of Section 11.3 on decomposing a vector relative to a nilpotent Lie algebra of endomorphism to find a weight space decomposition of g relative to $\mathrm{ad}_g h$.

2. Root Spaces of Split Semisimple Lie Algebras

In this section g will always denote a split semisimple Lie algebra over K and h will always denote a fixed, but arbitrary, Cartan subalgebra of g. Since $\mathrm{ad}_g h = \{\mathrm{ad}\, H : g \to g : H \in h\}$ is a nilpotent algebra of split linear transformations of g we know that g is a vector space direct sum of the weight spaces of the weight functions of $\mathrm{ad}_g h$.

Definition 13.6 The weight functions of $\mathrm{ad}_g h$ are called the **roots** of g and their weight spaces are called the **root spaces** of g; note Definition 10.15.

Remarks (1) Strictly speaking the weight spaces should be called the "root spaces of g with respect to h," but we will see later than the structure of root spaces does not depend on the choice of h.

(2) Using the previous notation for weight spaces $g(\alpha)$ will denote the root space of the root α. Proposition 13.4 implies that $g(0) = h$. It will be very useful to express g as the vector space direct sum

$$g = h + \sum_{\alpha \in \mathcal{R}} g(\alpha),$$

where $\mathcal{R}$ denotes the nonzero roots of g. As in Section 13.1 we obtain for α, β roots of h that

$$[g(\alpha)g(\beta)] \subset \begin{cases} g(\alpha+\beta) & \text{if } \alpha+\beta \text{ is a root,} \\ \{0\} & \text{otherwise.} \end{cases}$$

Proposition 13.7 Let g be a split semisimple Lie algebra over K, h a Cartan subalgebra of g, and as usual let $\mathrm{Kill}(X, Y)$ denote the Killing form of g.

(a) If α, β are roots of g with $\alpha + \beta \neq 0$ and $X \in g(\alpha)$, $Y \in g(\beta)$, then $\mathrm{Kill}(X, Y) = 0$.

(b) If $\alpha \in \mathcal{R}$, then $-\alpha \in \mathcal{R}$ and $\dim(g(-\alpha)) = \dim(g(\alpha))$.

(c) $\mathrm{Kill}(X, Y)$ restricted to h is nondegenerate.

(d) $[hh] = \{0\}$.

PROOF (a) Let $T = \operatorname{ad} X \operatorname{ad} Y$ and notice that $T^k(g(\gamma)) \subseteq g(\gamma + k(\alpha + \beta))$ for those positive integers k for which $\gamma + k(\alpha + \beta)$ is a root of g. Since the number of roots are finite we have that, for some large enough $k, \gamma + k(\alpha + \beta)$ will not be a root of g and $T^k(g(\gamma)) = \{0\}$. Since g is a direct sum of its root spaces, this implies that T is a nilpotent linear transformation so that Kill $(X, Y) = \operatorname{trace} T = 0$. Notice that we have included the possibility that α or β is the zero root and have used only the fact that $\alpha + \beta \neq 0$.

(b) Suppose $\alpha \in \mathscr{R}$ but $-\alpha \notin \mathscr{R}$. Then from (a) it follows that $\operatorname{Kill}(g(\alpha), g(\beta)) = 0$ for all roots β of g and $\operatorname{Kill}(g(\alpha), g) = 0$ which contradicts the nondegeneracy of $\operatorname{Kill}(X, Y)$ on g. In fact, general arguments about nondegenerate bilinear forms allow us to conclude that $g(\alpha)$ and $g(-\alpha)$ have bases $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$, respectively, with $\operatorname{Kill}(X_i, X_j) = \operatorname{Kill}(Y_i, Y_j) = 0$ for all i, j and $\operatorname{Kill}(X_i, Y_j) = \delta_{ij}$ which is the Kronecker δ-symbol. In particular, the dimensions of $g(\alpha)$ and $g(-\alpha)$ are the same.

(c) $0 \neq H_1 \in h$ implies $\operatorname{Kill}(H_1, \sum_{\alpha \in \mathscr{R}} g(\alpha)) = 0$. If also $\operatorname{Kill}(H_1, h) = 0$, then $\operatorname{Kill}(H_1, g) = 0$ contradicts the nondegeneracy of $\operatorname{Kill}(X, Y)$; so there is some $H_2 \in h$ with $\operatorname{Kill}(H_1, H_2) \neq 0$.

(d) $\operatorname{ad}_g h$ is a nilpotent and therefore a solvable Lie subalgebra of $gl(g)$. By Cartan's criterion for solvability we have

$$\operatorname{Kill}(H_1, [H_2 H_3]) = \operatorname{trace}(\operatorname{ad} H_1)(\operatorname{ad}[H_2 H_3]) = 0$$

for all $H_1, H_2, H_3 \in h$; that is $\operatorname{Kill}(h, [hh]) = 0$. Now (c) implies $[hh] = 0$.

Definition 13.8 (a) Let h^* denote the **dual space** of h; that is, h^* is the set of linear functions from h into K. Notice that $\mathscr{R}$ is a finite subset of h^*.

(b) For each $\alpha \in \mathscr{R}$ let H_α denote the unique element of h such that $\operatorname{Kill}(H_\alpha, H) = \alpha(H)$ for all $H \in h$. It is clear that such an H_α exists because of an elementary result which states that the conclusion is true for any symmetric nondegenerate bilinear form on a vector space and any function in the dual space of the vector space.

Proposition 13.9 Let g be any split semisimple Lie algebra with the notation as before.

(a) Then $\operatorname{Kill}(H_1, H_2) = \sum_{\alpha \in \mathscr{R}} n_\alpha \alpha(H_1)\alpha(H_2)$ for all $H_1, H_2 \in h$ where n_α is the dimension of $g(\alpha)$.

(b) Then $\mathscr{R}$ spans h^*.

(c) If $X \in g(\alpha)$, $Y \in g(-\alpha)$, and $(\operatorname{ad} H)X = \alpha(H)X$ for all $H \in h$, then $[XY] = \operatorname{Kill}(X, Y)H_\alpha$.

PROOF (a) $\operatorname{Kill}(H_1, H_2) = \operatorname{trace}(\operatorname{ad} H_1 \operatorname{ad} H_2)$ and the formula follows by looking at the matrices for $\operatorname{ad} H_1$ and $\operatorname{ad} H_2$ in Jordan canonical form.

(b) Let $f \in h^*$. Then using the remarks in Definition 13.8, we have an $H_f \in h$ such that $f(H) = \mathrm{Kill}(H_f, H)$ for all $H \in h$. However, from (a), $f(H) = \sum n_\alpha \alpha(H_f)\alpha(H)$ so that $\mathscr{R}$ spans h^*.

(c) For any $H \in h$,

$$\begin{aligned}\mathrm{Kill}([XY], H) &= \mathrm{Kill}([HX], Y) \\ &= \alpha(H)\,\mathrm{Kill}(X, Y) = \mathrm{Kill}\big(\mathrm{Kill}(X, Y)H_\alpha, H\big)\end{aligned}$$

and the formula follows from the nondegeneracy of $\mathrm{Kill}(X, Y)$ on h.

Proposition 13.10 Let g be a split semisimple Lie algebra as before. Then the root spaces $g(\alpha)$ for $\alpha \in \mathscr{R}$ are not only weight spaces but are in fact characteristic spaces; that is, $[HX] = \alpha(H)X$ for any $H \in h$ and $X \in g(\alpha)$.

Proof Choose a basis for g so that the matrices for each ad H with $H \in h$ restricted to $g(\alpha)$ look like

$$\begin{bmatrix} \alpha(H) & & * \\ & \ddots & \\ 0 & & \alpha(H) \end{bmatrix}.$$

For a fixed $H \in h$ write ad $H = S + N$ where S is semisimple and N is nilpotent. Then $S(X) = \alpha(H)X$ for any $X \in g(\alpha)$ and since $[g(\alpha)g(\beta)] \subseteq g(\alpha + \beta)$ one finds that

$$S([XY]) = (\alpha + \beta)(H)[XY] = [\alpha(H)XY] + [X\beta(H)Y] = [S(X)\,Y] + [XS(Y)]$$

for all $X \in g(\alpha)$ and $Y \in g(\beta)$. This shows that S is a derivation of g and from Section 12.3, $S = \mathrm{ad}\, Z$ for some $Z \in g$. By noticing that $S = \mathrm{ad}\, Z$ consist of the diagonal part of the matrix for ad H we find that $S = \mathrm{ad}\, Z$ commutes with all the matrices of $\mathrm{ad}_g\, h$ and so $[Zh] = \{0\}$ and $Z \in N(h) = h$. We must show that $S = \mathrm{ad}\, Z = \mathrm{ad}\, H$; that is, $N = \mathrm{ad}(H - Z)$ is 0. The matrix for $\mathrm{ad}(H - Z)$ is strictly upper triangular so $\mathrm{Kill}(H_1, H - Z) = \mathrm{trace}\big(\mathrm{ad}\, H_1\, \mathrm{ad}(H - Z)\big) = 0$ for all $H_1 \in h$, and $H - Z = 0$ by the nondegeneracy of $\mathrm{Kill}(X, Y)$ restricted to h. Thus $\mathrm{ad}(H - Z) = 0$ as required.

Remark (3) Notice that Proposition 13.10 implies that the formula in Proposition 13.9 (c) holds for all $X \in g(\alpha)$. This formula and some facts about representations of $sl(2, K)$ will be used in the next section to prove some further properties of root spaces. First, however, this section will be concluded with some examples.

Examples (1) Let $g = sl(n, K)$ and let E_{ij} be as usual, the matrices in $gl(n, K)$ with 1 in the ith row and jth column and zeros elsewhere. We first compute the Killing form for $sl(n, K)$. Since I commutes with everything in $gl(n, K)$ and $gl(n, K) = KI + sl(n, K)$, the Killing form of $sl(n, K)$ is equal to the Killing form of $gl(n, K)$ restricted to $sl(n, K)$. Thus for any $A = [a_{ij}]$, $B = [b_{ij}] \in g = sl(n, K)$, $\text{Kill}(A, B) = \text{trace}(\text{ad } A \text{ ad } B)$ where the trace is computed by using the E_{ij}'s as a basis for $gl(n, K)$.

$$\begin{aligned}(\text{ad } A \text{ ad } B)E_{ij} &= (\text{ad } A) \sum_{k=1}^{n} (b_{ki} E_{kj} - b_{jk} E_{ik}) \\ &= \sum_{m=1}^{n} \sum_{k=1}^{n} (a_{mk} b_{ki} E_{mj} - a_{jm} b_{ki} E_{km} - a_{mi} b_{jk} E_{mk} + a_{km} b_{jk} E_{im}).\end{aligned}$$

This contributes $\sum_{k=1}^{n} (a_{ik} b_{ki} + a_{kj} b_{jk}) - a_{jj} b_{ii} - a_{ii} b_{jj}$ to the trace and if we sum over all $1 \le i, j \le n$ we find

$$\text{Kill}(A, B) = 2n \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji} - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ii} b_{jj} = 2n(\text{trace}(AB))$$

since $\text{trace}(A) = \sum_{i=1}^{n} a_{ii} = 0$; note exercise (3), Section 12.2.

Notice that $\text{Kill}(A, B) = 2n(\text{trace}(AB))$ is nondegenerate which shows that $sl(n, K)$ is semisimple.

Let $h = \{\sum_{i=1}^{n} a_{ii} E_{ii} : \sum_{i=1}^{n} a_{ii} = 0\}$ which is the set of all diagonal matrices of $sl(n, K)$. It is easy to check that h is a Cartan subalgebra of $g = sl(n, K)$ and so g has rank $n - 1$. The nonzero roots of g are $\mathscr{R} = \{\alpha_{jk} : j \ne k,\ 1 \le j, k \le n\}$ where

$$\alpha_{jk}\left(\sum_{i=1}^{n} a_{ii} E_{ii}\right) = a_{jj} - a_{kk}$$

and the corresponding root spaces are

$$g(\alpha_{jk}) = KE_{jk}.$$

Notice that $-\alpha_{jk} = \alpha_{kj}$ and that each $g(\alpha_{jk})$ is one-dimensional. If we define $H_{jk} = (E_{jj} - E_{kk})/2n$, then, for any $H = \sum_{i=1}^{n} a_{ii} E_{ii} \in h$, $\text{Kill}(H_{jk}, H) = a_{jj} - a_{kk}$ so $H_{\alpha_{jk}} = H_{jk}$ where $H_{\alpha_{jk}}$ is given in Definition 13.8. It is also very easy to check that the conclusions of Proposition 13.9 also hold for this algebra. Thus $sl(n, K)$ will be an important example for later results and the above formulas will be used at that time.

(2) Consider the derivation algebra $\mathscr{D}(\mathscr{C})$ of a split Cayley algebra $\mathscr{C}$ over K as described in example (3) at the end of Section 9.6. Thus $\mathscr{D}(\mathscr{C}) = \{D(A, x, y) : A \in sl(3, K) \text{ and } x, y \in \mathscr{V} = K^3\}$, where $\mathscr{V}$ is the three-dimensional vector space of column vectors over K. The action of the derivations on $\mathscr{C}$ can be found in Chapter 9. However, the Lie multiplication

in $\mathscr{D}(\mathscr{C})$ is of more interest to us now and for convenience will be displayed once more.

$$[D(A_1, x_1, y_1), D(A_2, x_2, y_2)] = D(A_3, x_3, y_3)$$

where

$$\begin{aligned} A_3 &= [A_1, A_2] + 3x_1 y_2{}^t - 3x_2 y_1{}^t + ((x_2, y_1) - (x_1, y_2))I, \\ x_3 &= A_1 x_2 - A_2 x_1 - 2y_1 \times y_2 \\ y_3 &= -A_1{}^t y_2 + A_2{}^t y_1 + 2x_1 \times x_2 . \end{aligned}$$

The Killing form for $\mathscr{D}(\mathscr{C})$ can be computed as in the previous example using any convenient basis for $\mathscr{D}(\mathscr{C})$. One finds that

$$\operatorname{Kill}(D(A_1, x_1, y_1), D(A_2, x_2, y_2)) = 8 \operatorname{trace}(A_1 A_2) + 24(x_1, y_2) + 24(x_2, y_1)$$

so $\mathscr{D}(\mathscr{C})$ is semisimple since the Killing form is clearly nondegenerate.

To examine some of the other concepts of this chapter set $g = \mathscr{D}(\mathscr{C})$ and $h = \{D(H, 0, 0) : H \in sl(3, K), H \text{ diagonal}\}$. For $D(H, 0, 0) \in h$ the multiplication simplifies to

$$[D(H, 0, 0), D(A, x, y)] = D([H, A], Hx, -Hy).$$

It is now easy to check that h is Abelian and self-normalizing and so is a Cartan subalgebra of g, and g has rank 2.

Let

$$E_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad E_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad E_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

be the usual basis for $\mathscr{V} = K^3$ and let E_{ij}, $1 \le i, j \le 3$, be as in the previous example. Set $H = \sum_{i=1}^3 a_{ii} E_{ii} \in sl(3, K)$ so that $a_{11} + a_{22} + a_{33} = 0$. The roots and root spaces for g are now given in Table 13.1.

Notice that after we discover that g is a direct sum of its roots spaces we then know that g is in fact a split Lie algebra or at least that all the characteristic roots of ad H for $H \in h$ are in K. It will become apparent in the next section

TABLE 13.1

i	$\alpha_i(H)$	$g(\alpha_i)$	$g(-\alpha_i)$
1	$a_{11} - a_{22}$	$KD(E_{12}, 0, 0)$	$KD(E_{21}, 0, 0)$
2	$a_{11} - a_{33}$	$KD(E_{13}, 0, 0)$	$KD(E_{31}, 0, 0)$
3	$a_{22} - a_{33}$	$KD(E_{23}, 0, 0)$	$KD(E_{32}, 0, 0)$
4	a_{11}	$KD(0, E_1, 0)$	$KD(0, 0, E_1)$
5	a_{22}	$KD(0, E_2, 0)$	$KD(0, 0, E_2)$
6	a_{33}	$KD(0, E_3, 0)$	$KD(0, 0, E_3)$

that this implies that g is split. Notice that $\mathscr{C}$ being a split Cayley algebra implies that $\mathscr{D}(\mathscr{C})$ is a split semisimple Lie algebra even though the definition of a split Cayley algebra did not even involve characteristic roots of a linear transformation. It is in fact known that for any Cayley algebra $\mathscr{C}$, $\mathscr{D}(\mathscr{C})$ is split if and only if $\mathscr{C}$ is split as a composition algebra. The following exercises are related to the examples in Section 2.

Exercises (1) For any integer $n \geq 1$ let

$$J_n = \begin{bmatrix} 1 & 0 \cdots 0 & \\ 0 & O_n & I_n \\ \vdots & & \\ 0 & I_n & O_n \end{bmatrix}$$

be a $(2n+1) \times (2n+1)$ matrix where O_n and I_n are, respectively, $n \times n$ zero and $n \times n$ identity matrices. Let $\mathscr{B}(n, K)$ be the Lie algebra of all $(2n+1) \times (2n+1)$ matrices A over K such that $J_n A^t J_n = -A$. The notation for $\mathscr{B}(n, K)$ is chosen for historical reasons and will be explained in the next chapter. Verify the following facts about $\mathscr{B}(n, K)$:

(i) $\mathscr{B}(n, K)$ is just the set of matrices

$$\begin{bmatrix} 0 & u & v \\ -v^t & R & S \\ -u^t & T & -R^t \end{bmatrix}$$

where u, v are $n \times 1$ matrices, R, S, T are $n \times n$ matrices, $S^t = -S$, $T^t = -T$. The dimension of $\mathscr{B}(n, K)$ is $2n^2 + n$.

(ii) $\text{Kill}(A, B) = (2n-1)\,\text{trace}\, AB$ for any $A, B \in \mathscr{B}(n, K)$ and so $\mathscr{B}(n, K)$ is a semisimple algebra.

(iii) The subalgebra h of $\mathscr{B}(n, K)$ of diagonal matrices in $\mathscr{B}(n, K)$ is a Cartan subalgebra. The rank of $\mathscr{B}(n, K)$ is n.

(iv) Let $H = \sum_{i=1}^n a_{i+1, i+1}(E_{i+1, i+1} - E_{i+n+1, i+n+1})$ be an arbitrary element of h. The roots and root spaces of $g = \mathscr{B}(n, K)$ are described in Table 13.2.

TABLE 13.2

Root λ	$\lambda(H)$	Basis for $g(\lambda)$
$\alpha_{i,j}$, $1 \leq i, j \leq n, i \neq j$	$a_{i+1,i+1} - a_{j+1,j+1}$	$E_{i+1,j+1} - E_{j+n+1,i+n+1}$
$\beta_{i,j}$, $1 \leq i < j \leq n$	$a_{i+1,i+1} + a_{j+1,j+1}$	$E_{i+1,j+n+1} - E_{j+1,i+n+1}$
$\gamma_{i,j}$, $1 \leq i < j \leq n$	$-a_{i+1,i+1} - a_{j+1,j+1}$	$E_{i+n+1,i+1} - E_{j+n+1,i+1}$
δ_i, $1 \leq i \leq n$	$-a_{i+1,i+1}$	$E_{1,i+1} - E_{i+n+1,1}$
ε_i, $1 \leq i \leq n$	$a_{i+1,i+1}$	$E_{1,i+n+1} - E_{i+1,1}$

(2) For any $n \geq 1$, let

$$L_n = \begin{bmatrix} O_n & I_n \\ I_n & O_n \end{bmatrix}$$

be the $2n \times 2n$ matrix where O_n and I_n are as in the previous exercise. Let $\mathscr{D}(n, K)$ be the Lie algebra of all $2n \times 2n$ matrices A over K such that $L_n A^t L_n = -A$. Verify the following facts about $\mathscr{D}(n, K)$:

(i) $\mathscr{D}(n, K)$ is just the set of matrices

$$\begin{bmatrix} R & S \\ T & -R^t \end{bmatrix}$$

where R, S, T are $n \times n$ matrices and $S^t = -S$, $T^t = -T$. The dimension of $\mathscr{D}(n, K)$ is $2n^2 - n$.

(ii) $\operatorname{Kill}(A, B) = (2n - 2) \operatorname{trace} AB$ for any $A, B \in \mathscr{D}(n, K)$, so $\mathscr{D}(n, K)$ is semisimple.

(iii) The subalgebra of $\mathscr{D}(n, K)$ of diagonal matrices is a Cartan subalgebra of $\mathscr{D}(n, K)$. The rank of $\mathscr{D}(n, K)$ is n.

(iv) Use (iv) of the previous exercise to produce a table of roots and root spaces of $\mathscr{D}(n, K)$.

(3) For any $n \geq 1$, let

$$M_n = \begin{bmatrix} O_n & I_n \\ -I_n & O_n \end{bmatrix}$$

be the $2n \times 2n$ matrix where O_n and I_n are as before. Let $\mathscr{C}(n, K)$ be the Lie algebra of all $2n \times 2n$ matrices A over K such that $M_n A^t M_n^{-1} = -A$. Verify the following facts about $\mathscr{C}(n, K)$:

(i) $\mathscr{C}(n, K)$ is just the set of matrices

$$\begin{bmatrix} R & S \\ T & -R^t \end{bmatrix}$$

where R, S, T are $n \times n$ matrices and $S^t = S$, $T^t = T$. The dimension of $\mathscr{C}(n, K)$ is $2 + n^2 n$.

(ii) $\operatorname{Kill}(AB) = (2n + 2) \operatorname{trace} AB$ for any $A, B \in \mathscr{C}(n, K)$ so $\mathscr{C}(n, K)$ is semisimple.

(iii) The subalgebra of $\mathscr{C}(n, K)$ of diagonal matrices is a Cartan subalgebra. The rank of $\mathscr{C}(n, K)$ is n.

(iv) Let $H = \sum_{i=1}^{n} a_{i,i}(E_{i,i} - E_{i+n,i+n})$ be an arbitrary element of h. The roots and root spaces of $g = \mathscr{C}(n, K)$ are described in Table 13.3.

REMARK (4) The examples and exercises above include all but four of the simple split Lie algebras over a fixed field K of characteristic 0. The proof of this fact (which follows in the next chapter) includes many algebraic computations, but the reader should be assured that the above matrix computa-

TABLE 13.3

Root λ	$\lambda(H)$	Basis for $g(\lambda)$
$\alpha_{i,j}$, $1 \le i, j \le n, i \ne j$	$a_{i,i} - a_{j,j}$	$E_{i,j} - E_{j+n,i+n}$
$\beta_{i,j}$, $1 \le i \le j \le n$	$a_{i,i} + a_{j,j}$	$E_{i,j+n} + E_{j,i+n}$
$\gamma_{i,j}$, $1 \le i \le j \le n$	$-a_{i,i} - a_{j,j}$	$E_{i+n,j} + E_{j+n,i}$
$2\delta_i$, $1 \le i \le n$	$-2a_{ii}$	$E_{i+n,i}$
$2\varepsilon_i$, $1 \le i \le n$	$2a_{ii}$	$E_{i,i+n}$

tions comprise the bulk of the tiresome details needed in describing the simple Lie algebras.

(5) The algebras $\mathscr{B}(n, K)$, $\mathscr{C}(n, K)$, and $\mathscr{D}(n, K)$ are often described as the set of linear transformations that are skew with respect to certain bilinear forms. Thus let J be one of the matrices J_n, L_n, or M_n appearing in the exercises and let X and Y be column vectors of the appropriate size so that $B(X, Y) = X^t JY$ defines a bilinear form. Notice that $B(X, Y)$ is nondegenerate in all three cases, that it is symmetric in the first two cases, and that it is antisymmetric in the third case. For a linear transformation A on the given vector space of column vectors let A^* be the unique linear transformation such that $B(AX, Y) = B(X, A^*Y)$ for all X and Y, one can easily check that $A^* = J^{-1}A^tJ$. Then g, the Lie algebra of all linear transformations with $A^* = -A$, is isomorphic to the Lie algebras of matrices that were described in the exercises.

(6) Summarizing some of the results for a split semisimple Lie algebra g we have

$$g = h + \sum g(\alpha) = h + \sum g(\alpha) + g(-\alpha),$$

where h is a Cartan subalgebra of g and α varies over $\mathscr{R}$. The second sum is not direct, since we are duplicating some root spaces. In the next section we will show $h = \sum KH_\alpha$ for $H_\alpha \in h$ of Definition 13.8 and α varies over $\mathscr{R}$. Thus

$$g = \sum KH_\alpha + g(\alpha) + g(-\alpha),$$

and we shall show $KH_\alpha + g(\alpha) + g(-\alpha)$ is isomorphic to $sl(2, K)$; that is, g is built up of three-dimensional Lie subalgebras which are isomorphic to $sl(2, K)$.

3. Irreducible Representations of $sl(2, K)$

The following basic theorem completely characterizes the finite-dimensional irreducible representations of $sl(2, K)$. This theorem will be used to obtain further properties of roots of split semisimple Lie algebras.

Theorem 13.11 Let g be an abstract Lie algebra over K with basis H, E, F where $[HE] = 2E$, $[HF] = -2F$, and $[EF] = H$ so that g is isomorphic to $sl(2, K)$. For any positive integer $n \geq 2$ define a linear map $\rho : g \to gl(n, K)$ by setting

$$\rho(H) = \begin{bmatrix} n-1 & 0 & 0 & & & 0 \\ 0 & n-3 & 0 & & & \\ 0 & 0 & n-5 & & & \\ & & & \ddots & & \\ & 0 & & & 3-n & 0 \\ & & & & 0 & 1-n \end{bmatrix},$$

$$\rho(E) = \begin{bmatrix} 0 & n-1 & 0 & & & & \\ 0 & 0 & n-2 & & & 0 & \\ 0 & 0 & 0 & & & & \\ & & & \ddots & & & \\ & & & & 0 & 2 & 0 \\ & 0 & & & 0 & 0 & 1 \\ & & & & 0 & 0 & 0 \end{bmatrix},$$

$$\rho(F) = \begin{bmatrix} 0 & 0 & 0 & & & & \\ 1 & 0 & 0 & & & 0 & \\ 0 & 2 & 0 & & & & \\ & & & \ddots & & & \\ & & & & 0 & 0 & 0 \\ & 0 & & & n-2 & 0 & 0 \\ & & & & 0 & n-1 & 0 \end{bmatrix}.$$

Then ρ is an irreducible representation of g. Conversely, given any finite-dimensional irreducible representation $\rho : g \to gl(V)$, there exists a basis for V so that the matrices for $\rho(H)$, $\rho(E)$, and $\rho(F)$ are precisely those above with n being the dimension of V.

PROOF A straightforward matrix computation will show that $[\rho(H), \rho(E)] = 2\rho(E)$, $[\rho(H), \rho(F)] = -2\rho(F)$, and $[\rho(E), \rho(F)] = \rho(H)$. Then ρ is a representation of g. To show that $\rho(g)$ acts irreducibly on the n-dimensional vector space V of column vectors we first consider the usual basis for V. Thus X_k is the vector with a 1 in the kth row and zeros elsewhere. Let $X = \sum_{k=1} x_k X_k$ be any nonzero vector in V. We must show that any subspace of V containing X which is invariant under $\rho(g)$ must be equal to V itself.

Suppose $x_k = 0$ for $1 \leq k < m \leq n$ but $x_m \neq 0$. Then notice if $m \neq n$ that

$$\rho(F)X = \sum_{k=1}^{n-1} kx_k X_{k+1} = \sum_{k=m}^{n-1} kx_k X_{k+1}$$

so $\rho(F)^{n-m}X = m(m+1)\cdots(n-1)x_m X_n$. Thus any nontrivial invariant subspace of V contains X_n. However, $\rho(E)^k X_n = k!\, X_{n-k}$ so any invariant subspace contains all the vectors in the basis of V and must be V itself. We have shown that ρ is an irreducible representation of g and will now show that these are the only possible finite-dimensional irreducible representations.

Conversely, given ρ, we first claim that there exists a $0 \neq X \in V$ with $\rho(H)X = aX$ for some $a \neq 0$ and $\rho(E)X = 0$. So first assume K is algebraically closed so that $\rho(H)$ has a characteristic vector, say $\rho(H)Y = bY$. We will see shortly that b must be an integer so at that point we no longer need the assumption of algebraic closure. If $\rho(E)Y = 0$, we are finished. Otherwise let $Y_2 = \rho(E)Y$ and notice that

$$\begin{aligned} \rho(H)Y_2 &= \rho(H)\rho(E)Y \\ &= \big(\rho([HE]) + \rho(E)\rho(H)\big)Y \\ &= 2\rho(E)Y + b\rho(E)Y = (b+2)Y_2. \end{aligned}$$

If $\rho(E)Y_2 \neq 0$, set $Y_3 = \rho(E)Y_2$ and compute that $\rho(H)Y_3 = (b+4)Y_3$. In this way we obtain a sequence $Y, Y_2, Y_3, \ldots$ of characteristic vectors of $\rho(H)$ each with a different characteristic value so the vectors of this sequence are linearly independent. Since V is finite-dimensional there can be only a finite number of vectors in this sequence; that is, $\rho(E)Y_k = 0$ for some k and Y_k is the vector desired.

Let X_1 be a vector as constructed in the previous paragraph so $\rho(H)X_1 = aX_1$ and $\rho(E)X_1 = 0$. For each positive integer k define $X_k = \rho(F)^{k-1}X_1/(k-1)! = \rho(F)X_{k-1}/(k-1)$. As in the previous paragraph it is easy to verify that $\rho(H)X_k = (a-2k+2)X_k$ so only a finite number of the X_k's are nonzero. Choose n so that X_n is the last nonzero vector in the sequence. The $X_1, X_2, \ldots, X_n$ form a basis for V since the following three formulas show that the subspace spanned by these vectors is invariant under $\rho(H)$, $\rho(E)$, $\rho(F)$

$$\begin{aligned} &\rho(H)X_k = (a-2k+2)X_k, \\ &\rho(F)X_k = kX_{k+1} && \text{for } 1 \leq k < n, \quad \rho(F)X_n = 0, \\ &\rho(E)X_k = (a-k+2)X_{k-1} && \text{for } 2 \leq k \leq n, \quad \rho(E)X_1 = 0. \end{aligned}$$

We must still prove the last formula. However, first notice that the formulas precisely coincide with matrix entries we are interested in if $a = n - 1$.

To prove the formula for $\rho(E)$ notice that

$$\begin{aligned} \rho(E)X_2 &= \rho(E)\rho(F)X_1 \\ &= \big(\rho(H) + \rho(F)\rho(E)\big)X_1 = aX_1. \end{aligned}$$

Now proceed by mathematical induction and assume the result for $k-1$. Then

$$\begin{aligned}(k-1)\rho(E)X_k &= \rho(E)\rho(F)X_{k-1} \\ &= \big(\rho(H)+\rho(F)\rho(E)\big)X_{k-1} \\ &= (a-2k+4)X_{k-1} + (k-2)(a-k+3)X_{k-1} \\ &= (k-1)(a-k+2)X_{k-1}.\end{aligned}$$

The desired formula now follows. Finally to show that $a = n-1$ notice that the matrix for $\rho(H) = [\rho(E), \rho(H)]$ is of trace 0 so $a + (a-2) + (a-4) + \cdots + (a-2n+2) = 0 = na - n(n-1)$. This completes the proof of the theorem.

Example (1) If $n = 2$ in the theorem, one obtains the usual matrices for $sl(2, K)$.

Exercise (1) The algebra $\mathscr{B}(1, K)$ described in the first exercise of the previous section is a three-dimensional algebra of 3×3 matrices and is isomorphic to $sl(2, K)$. An explicit isomorphism from the algebra g described in Theorem 13.11 to $\mathscr{B}(1, K)$ is indicated by $H \to H'$, $E \to E'$, $F \to F'$, where

$$H' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \qquad E' = \begin{bmatrix} 0 & 0 & -2 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad F' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

Show that the matrices of $\mathscr{B}(1, K)$ describe an irreducible representation of $sl(2,K)$ of degree 3. Find a nonsingular 3×3 matrix P so that $PH'P^{-1} = \rho(H)$, $PE'P^{-1} = \rho(E)$, $PF'P^{-1} = \rho(F)$, where $\rho(H)$, $\rho(E)$, $\rho(F)$ are as in Theorem 13.11 with $n = 3$.

REMARK (1) For the remainder of this section we will return to investigating split semisimple Lie algebras. The notation will be as in the previous section so g is a split semisimple Lie algebra over a field K of characteristic 0 and h is a Cartan subalgebra of g. We know that $g = h + \sum_{\alpha \in \mathscr{R}} g(\alpha)$ as a vector space direct sum where $\mathscr{R}$ denotes the nonzero roots of g and each root space $g(\alpha)$ is a characteristic space. We wish to find a subalgebra of g isomorphic to $sl(2, K)$. To find such a subalgebra we can use Proposition 13.9(c) that says that for any $X \in g(\alpha)$ and $Y \in g(-\alpha)$ we have $[XY] = \mathrm{Kill}(X, Y)H_\alpha$ where $\mathrm{Kill}(H, H_\alpha) = \alpha(H)$ for all $H \in h$.

Proposition 13.12 With the notation as above we have:

(a) $\alpha(H_\alpha) = \mathrm{Kill}(H_\alpha, H_\alpha) \neq 0$ for each $\alpha \in \mathscr{R}$;

(b) the dimension of $g(\alpha)$ is 1 for each $\alpha \in \mathscr{R}$;

(c) if we define $h(\alpha) = [g(\alpha)g(-\alpha)] = KH_\alpha$, then $h(\alpha) + g(\alpha) + g(-\alpha)$ is a subalgebra of g isomorphic to $sl(2, K)$.

PROOF (a) Suppose for a fixed $\alpha \in \mathscr{R}$ we have $\alpha(H_\alpha) = 0$. Then choose any $X \in g(\alpha)$, $Y \in g(-\alpha)$ so that $\text{Kill}(X, Y) = 1$. Notice $[H_\alpha X] = \alpha(H_\alpha)X = 0$, $[H_\alpha Y] = 0$, $[XY] = H_\alpha$. Thus H_α, X, Y span a three-dimensional nilpotent subalgebra of g and ad H_α, ad X, ad Y span a three-dimensional nilpotent (and solvable) subalgebra of $gl(g)$. Since ad $H_\alpha = [\text{ad } X, \text{ad } Y]$ and $[\text{ad } H_\alpha, \text{ad } X] = 0$, we have by Lemma 12.34 that ad H_α is a nilpotent linear transformation. By Proposition 13.10, ad H_α is a diagonalizable linear transformation. A linear transformation which is both nilpotent and diagonalizable must be 0 so ad $H_\alpha = 0$. This clearly contradicts the semisimplicity of g.

(b) and (c) For a fixed $\alpha \in \mathscr{R}$ choose any $X \in g(\alpha)$ and $Y \in g(-\alpha)$ so that $\text{Kill}(X, Y) = 2/\alpha(H_\alpha)$ and set $H = 2H_\alpha/\alpha(H_\alpha)$. Then $[HX] = \alpha(H)X = 2X$, $[HY] = -2Y$, and $[XY] = \text{Kill}(X, Y)H_\alpha = H$ so $\bar{g}$, the subspace of g spanned by H, X, Y, is a subalgebra of g isomorphic to $sl(2, K)$. To complete the proof we must show that $g(\alpha)$ is one dimensional or equivalently that $g(-\alpha)$ is one dimensional.

Suppose the dimension of $g(-\alpha)$ is greater than 1, then choose some $0 \neq Z \in g(-\alpha)$ with $\text{Kill}(X, Z) = 0$; note the proof of Proposition 13.7(b). Then

$$(\text{ad } X)Z = [XZ] = \text{Kill}(X, Z)H_\alpha = 0, \qquad (\text{ad } H)Z = [HZ] = -\alpha(H)Z = -2Z.$$

This is precisely the situation described in the last part of the proof of Theorem 13.11, except for an unfortunate clash of notation. We could follow the computations there to show that $Z_1 = Z$, $Z_2 = (\text{ad } Y)Z$, $Z_3 = (\text{ad } Y)^2 Z/2!, \ldots,$ $Z_n = (\text{ad } Y)^{n-1}Z/(n-1)!$ forms a basis for a subspace W of g invariant under ad X, ad Y, and ad H. Furthermore we can conclude from previous computations that ad H restricted to W has a matrix

$$\lambda(H) = \begin{bmatrix} -2 & 0 & 0 & & \\ 0 & -4 & 0 & & 0 \\ 0 & 0 & -6 & & \\ & & & \ddots & \\ & 0 & & & -2n \end{bmatrix}$$

where λ = ad restricted to W. On the other hand $\lambda(H) = [\lambda(X), \lambda(Y)]$ on W, so that trace $\lambda(H) = 0$. This is a contradiction which shows that $g(-\alpha)$ is one-dimensional.

Definition 13.13 For each $\alpha \in \mathscr{R}$, let

$$H_\alpha' = 2H_\alpha/\alpha(H_\alpha),$$

where, as before, H_α is the unique element of h such that $\text{Kill}(H_\alpha, H) = \alpha(H)$ for all $H \in h$.

REMARK (2) Clearly $[H_\alpha'X] = 2X$ for any $X \in g(\alpha)$ and $[H_\alpha'Y] = -2Y$ for any $Y \in g(-\alpha)$. Given any nonzero $X \in g(\alpha)$, there exists a unique $Y \in g(-\alpha)$ with $[XY] = H_\alpha'$ and then H_α', X, Y multiplies like the usual basis for $sl(2, K)$.

Definition 13.14 For any $\alpha, \beta \in \mathscr{R}$, define $\langle\alpha, \beta\rangle = \mathrm{Kill}(H_\alpha, H_\beta)$. Since $\mathscr{R}$ spans h^*, the dual space of h, this definition can be extended to a symmetric nondegenerate bilinear form $\langle\lambda, \mu\rangle$ on h^*.

REMARK (3) Definition 13.14 can be thought of as a simplification of notation. Thus $H_\alpha' = 2H_\alpha/\langle\alpha, \alpha\rangle$, and from Proposition 13.9

$$\langle\lambda, \mu\rangle = \sum_{\alpha \in \mathscr{R}} \langle\lambda, \alpha\rangle\langle\mu, \alpha\rangle$$

for any $\lambda, \mu \in \mathscr{R}$. Now if the base field K is either the real numbers or the rational numbers, then for any $\beta \in \mathscr{R}$

$$\langle\beta, \beta\rangle = \sum_{\alpha \in \mathscr{R}} \langle\beta, \alpha\rangle^2 > 0.$$

Thus $\langle\lambda, \mu\rangle$ is positive definite in this case.

We make a few more observations. For $\lambda \in h^*$, let $\lambda(H) = \mathrm{Kill}(H_\lambda, H)$. Then if $\lambda = \sum a_i \alpha_i$ by Proposition 13.9, we have $H_\lambda = \sum a_i H_{\alpha_i}$, for $\lambda(H) = \sum a_i \alpha_i(H) = \sum a_i \mathrm{Kill}(H_{\alpha_i}, H) = \mathrm{Kill}(\sum a_i H_{\alpha_i}, H)$ and the result follows.

Next note for $\lambda, \mu \in h^*$ that

$$\langle\lambda, \mu\rangle = \mathrm{Kill}(H_\lambda, H_\mu),$$

for writing $\lambda = \sum a_i \alpha_i$ and $\mu = \sum b_j \alpha_j$, we see $\langle\lambda, \mu\rangle = \sum a_i b_j \langle\alpha_i, \alpha_j\rangle = \sum a_i b_j \mathrm{Kill}(H_{\alpha_i}, H_{\alpha_j}) = \mathrm{Kill}(H_\lambda, H_\mu)$ using the preceding paragraph.

Proposition 13.15 If $\alpha, \beta \in \mathscr{R}$ and β is not a multiple of α, then $m = 2\langle\alpha, \beta\rangle/\langle\alpha, \alpha\rangle$ is an integer, and $\beta, \beta - \alpha, \beta - 2\alpha, \ldots, \beta - m\alpha$ are all in $\mathscr{R}$ if $m > 0$ and $\beta, \beta + \alpha, \beta + 2\alpha, \ldots, \beta - m\alpha$ are in $\mathscr{R}$ if $m < 0$.

PROOF Choose $X \in g(\alpha)$, $Y \in g(-\alpha)$ so that H_α', X, Y are a basis for a subalgebra of g isomorphic to $sl(2, K)$. Now ad H_α', ad X, and ad Y act on g, so these maps restricted to an irreducible submodule of g must have matrices like those of Theorem 13.11. Any root space of g must also be a weight space in an irreducible submodule of g; in particular, this is true for $g(\beta)$. For any $Z \in g(\beta)$

$$[H_\alpha'Z] = \beta(2H_\alpha/\langle\alpha, \alpha\rangle)Z = (2\langle\alpha, \beta\rangle/\langle\alpha, \alpha\rangle)Z = mZ.$$

Thus m is a characteristic value of ad H_α'. Thus by comparing with $\rho(H)$ in Theorem 13.11, m must be an integer. We are finished with the proof if $m = 0$.

If $m > 0$, then by again examining $\rho(H)$ we find that $m, m-2, m-4, \ldots, -m$ are all characteristic values of characteristic spaces of $\mathrm{ad}(H_\alpha')$ in the same irreducible subspace of g as $g(\beta)$, in fact Z, $(\mathrm{ad}\ Y)Z$, $(\mathrm{ad}\ Y)^2Z, \ldots, (\mathrm{ad}\ Y)^mZ$, form a basis for these spaces, and these elements lie in $g(\beta), g(\beta-\alpha), g(\beta-2\alpha), \ldots, g(\beta-m\alpha)$ as required. The same argument works for $m<0$ by using $\mathrm{ad}\ X$ instead of $\mathrm{ad}\ Y$. Also note exercise (4) of this section.

Proposition 13.16 Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathscr{R}$ be a basis of h^* over K and let V denote the vector space over the rational numbers Q spanned by $\alpha_1, \alpha_2, \ldots, \alpha_n$. If $\alpha, \beta \in \mathscr{R}$, then $\langle\alpha, \beta\rangle \in Q$ and thus $\langle\lambda, \mu\rangle \in Q$ for all $\lambda, \mu \in V$. Also $\langle\lambda, \mu\rangle$ is positive definite on V.

PROOF By remark (3), $\langle\beta, \beta\rangle = \sum_{\alpha\in\mathscr{R}} \langle\beta, \alpha\rangle^2 > 0$ and therefore

$$4/\langle\beta, \beta\rangle = 4\langle\beta, \beta\rangle/\langle\beta, \beta\rangle^2 = \sum_{\alpha\in\mathscr{R}} 4\langle\beta, \alpha\rangle^2/\langle\beta, \beta\rangle^2$$

which is a sum of integers squared, using Proposition 13.15. Thus $\langle\beta, \beta\rangle \in Q$ and since $2\langle\alpha, \beta\rangle/\langle\beta, \beta\rangle$ is an integer, $\langle\alpha, \beta\rangle \in Q$. Since $\lambda, \mu \in V$ are rational linear combinations of elements of $\mathscr{R}$, $\langle\lambda, \mu\rangle \in Q$ also.

Now $\langle\lambda, \mu\rangle$ is positive definite on V as follows. From remark (3) we see for $\lambda \in V$, with $\lambda(H) = \mathrm{Kill}(H_\lambda, H)$ for $H \in h$, that

$$\begin{aligned}\langle\lambda, \lambda\rangle &= \mathrm{Kill}(H_\lambda, H_\lambda)\\ &= \sum n_\alpha \alpha(H_\lambda)\alpha(H_\lambda) = \sum \alpha(H_\lambda)^2\end{aligned}$$

using Proposition 13.9, and the dimension of $g(\alpha)$ is 1. Thus $\langle\lambda, \lambda\rangle = 0$ implies $\alpha(H_\lambda) = 0$ for all $\alpha \in \mathscr{R}$, which gives $H_\lambda = 0$ since $\mathscr{R}$ spans h^*. Thus $\lambda = 0$.

Proposition 13.17 If $\alpha, \beta \in \mathscr{R}$ and p, q are the largest nonnegative integers such that $\beta + p\alpha \in \mathscr{R}$ and $\beta - q\alpha \in \mathscr{R}$, then $\beta - q\alpha, \beta - (q-1)\alpha, \ldots, \beta + p\alpha$ are all roots of g and $2\langle\alpha, \beta\rangle/\langle\alpha, \alpha\rangle = q - p$.

PROOF Continue the notation in Proposition 13.15 so that $m = 2\langle\alpha, \beta\rangle/\langle\alpha, \alpha\rangle$. For $Z_1 \in g(\beta + p\alpha)$ and $Z_2 \in g(\beta - q\alpha)$, we have

$$\begin{aligned}(\mathrm{ad}\ H_\alpha')Z_1 &= (m+2p)Z_1, \qquad & (\mathrm{ad}\ X)Z_1 &= 0,\\ (\mathrm{ad}\ H_\alpha')Z_2 &= (m-2q)Z_2, \qquad & (\mathrm{ad}\ Y)Z_2 &= 0.\end{aligned}$$

Both Z_1 and Z_2 can be used to generate irreducible subspaces m_1 and m_2 under the action of the algebra spanned by $\mathrm{ad}\ H_\alpha$, $\mathrm{ad}\ X$, and $\mathrm{ad}\ Y$. Using irreducibility, these two subspaces must either coincide or have zero intersections. By checking the matrices of Theorem 13.11, we discover that the dimensions of m_1 and m_2 are $m + 2p + 1$ and $-(m - 2q) + 1$. The sum of the two dimensions is $2(p+q+1)$, so the spaces coincide and are in fact equal to $g(\beta - q\alpha) + g(\beta - (q-1)\alpha) + \cdots + g(\beta + p\alpha)$. Thus all $\beta + k\alpha$ are roots for $-q \le k \le p$ (with k integral). Finally, since the dimension of m_1 is $p+q+1$, we see $p+q+1 = m+2p+1 = 2q-m+1$ so that $m = q-p$ as required.

Exercises (2) Suppose r is the largest positive integer such that, for a given $\alpha \in \mathscr{R}$, we also have $r\alpha \in \mathscr{R}$. Let

$$k = g(-\alpha) + KH_\alpha' + g(\alpha) + g(2\alpha) + \cdots + g(r\alpha).$$

Show that k is a subalgebra of g which is invariant under ad H_α', ad X, and ad Y where $X \in g(\alpha)$, $Y \in g(-\alpha)$.

(3) By considering the trace of ad $H_\alpha' = [\text{ad } X, \text{ad } Y]$ restricted to k in exercise (2), show that $0, \alpha, -\alpha$ are the only integral multiples of α that are also roots of g.

(4) Show that if $\alpha \in \mathscr{R}$, $r \in K$, and $r\alpha \in \mathscr{R}$, then $r = 0$ or ± 1.

(5) Consider the Lie algebra $g = \mathscr{D}(\mathscr{C})$ of example (2) of the previous section. Letting $\beta = \alpha_4$ and $\alpha = \alpha_5$, verify directly the properties in the conclusion of Proposition 13.17.

(6) Given g an arbitrary semisimple split Lie algebra as before, show that if $\alpha, \beta, \alpha + \beta \in \mathscr{R}$ and $0 \ne X \in g(\alpha)$, $0 \ne Y \in g(\beta)$, then $[XY] \ne 0$.

Proposition 13.18 Consider a semisimple split Lie algebra g over a field K of characteristic 0 as before. If $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathscr{R}$ is a basis for h^* over K and if V is the vector space over the rational numbers Q spanned by $\alpha_1, \ldots, \alpha_n$, then $\mathscr{R} \subset V$.

PROOF Following the proof of Jacobson [1962], suppose $\beta = \sum_{i=1}^n t_i \alpha_i \in \mathscr{R}$. We must show that each $t_i \in Q$. Consider the system of equations

$$2\langle \beta, \alpha_j \rangle / \langle \alpha_j, \alpha_j \rangle = \sum_{i=1}^{n} t_i 2\langle \alpha_i, \alpha_j \rangle / \langle \alpha_j, \alpha_j \rangle$$

for $j = 1, 2, \ldots, n$. This system of equations in $t_1, t_2, \ldots, t_n$ has integral coefficients and will have a rational solution if that solution is unique. However, the uniqueness follows from the linear independence of the α_i's over the field K.

REMARK (4) Given g as in Proposition 3.18 and a fixed $\alpha \in \mathscr{R}$, define a nonsingular linear transformation $S_\alpha : V \to V$ by

$$S_\alpha(\lambda) = \lambda - 2(\langle \alpha, \lambda \rangle / \langle \alpha, \alpha \rangle)\alpha$$

for all $\lambda \in V$. Proposition 13.15 guarantees that $S_\alpha(\lambda) \in V$ and at the same time shows that $S_\alpha(\beta) \in \mathscr{R}$ for all $\beta \in \mathscr{R}$. Thus $S_\alpha(\mathscr{R}) = \mathscr{R}$; that is, S_α permutes the finite number of elements of $\mathscr{R}$. Notice also that $S_\alpha(\alpha) = -\alpha$ and $U = \{\lambda \in V \mid S_\alpha(\lambda) = \lambda\}$ is a subspace of V of dimension one less that the dimension of V. These will be properties of $\mathscr{R}$ and V that will be very important in the next chapter.

One can consider the group W of linear transformations on V generated by the S_α's as α varies over the elements of $\mathscr{R}$. This group is finite because it can also be thought of as a set of permutations on the finite set $\mathscr{R}$. Moreover this group actually only depends on the Lie algebra g and is called the **Weyl group of g.**

Exercise (7) Find the Weyl groups of $sl(2, K)$ and $sl(3, K)$.

CHAPTER 14

SIMPLE SPLIT LIE ALGEBRAS

In this chapter we examine the important results on roots of a semisimple split Lie algebra given in Chapter 13. This leads to the study of abstract root systems by means of Dynkin diagrams, and we classify the irreducible root systems. Conversely, in the second section we construct some models of simple Lie algebras corresponding to the irreducible root systems. These algebras consist of four classical matrix types which we have previously considered in examples and five exceptional types. In the third section we discuss the inner automorphisms of these simple algebras in terms of symmetries of the corresponding Dynkin diagram.

1. Root Systems

We now consider some of the previous results on roots and study abstract root systems. The Dynkin diagram of a root system is introduced, and by means of these diagrams we find the root systems which correspond to simple split Lie algebras.

Definition 14.1 Let V be a finite-dimensional vector space over Q, the field of rational numbers, and let (x, y) be a positive definite symmetric bilinear form on V.

(a) A finite subset $\mathscr{R}$ of nonzero vectors of V is called a **root system** in V if:

(i) $\mathscr{R}$ spans V,
(ii) $\alpha \in \mathscr{R}$ and $t\alpha \in \mathscr{R}$ with $t \in Q$, then $t = \pm 1$,

(iii) $\alpha, \beta \in \mathcal{R}$, then $2(\alpha, \beta)/(\alpha, \alpha)$ is an integer;
(iv) $\alpha, \beta \in \mathcal{R}$, then $\beta - 2[(\alpha, \beta)/(\alpha, \alpha)]\,\alpha \in \mathcal{R}$.

The elements of $\mathcal{R}$ are called **roots**.

(b) A root system is called **irreducible** if there is no proper subset $\mathcal{S} \subseteq \mathcal{R}$ such that $(\alpha, \beta) = 0$ for all $\alpha \in \mathcal{S}$ and $\beta \in \mathcal{R}$ but $\beta \notin \mathcal{S}$.

(c) For each $\alpha \in \mathcal{R}$ define the **α-symmetry** of V, $S_\alpha : V \to V$, by $S_\alpha(x) = x - 2[(\alpha, x)/(\alpha, \alpha)]\,\alpha$.

The **Weyl group** W of the root system $\mathcal{R}$ is the group generated by the S_α's as α varies over the roots of $\mathcal{R}$.

Example (1) Let g be a split semisimple Lie algebra over a field of characteristic 0, $\mathcal{R}$ is its nonzero roots, V as in Proposition 13.18, and (λ, μ) is the restriction of $\langle \lambda, \mu \rangle$ to V. Various propositions of Chapter 13 show us that $\mathcal{R}$ is in fact a root system in V.

REMARKS (1) Many authors give a less restrictive definition of root systems and then prove properties (i)–(iv) in Definition 14.1(a). Our definition was chosen to correspond with properties already proved in Chapter 13. Also authors use different names for irreducible systems such as **indecomposable systems**.

(2) Certain properties proved for roots in Chapter 13 are obvious from Definition 14.1. Thus $S_\alpha(\mathcal{R}) = \mathcal{R}$ for all $\alpha \in \mathcal{R}$; the elements of the Weyl group are completely determined by their action on $\mathcal{R}$; the Weyl group is finite; and if $\alpha \in \mathcal{R}$, then $-\alpha = S_\alpha(\alpha) \in \mathcal{R}$.

Proposition 14.2 If $\mathcal{R} \subset V$ is any root system, then there exist subspaces $V_i \subset V$ with $V = V_1 + V_2 + \cdots + V_m$ as a vector space direct sum, and $(x, y) = 0$ if $x \in V_i$, $y \in V_j$, $i \neq j$. Furthermore $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \cup \mathcal{R}_m$ as a disjoint union where $\mathcal{R}_i = \mathcal{R} \cap V_i$ is an irreducible root system in V_i. Thus to determine all root systems, we need only find the irreducible ones.

PROOF If $\mathcal{R} \subset V$ is not irreducible by definition, $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, a disjoint and orthogonal union. Now set V_i as the span of $\mathcal{R}_i$, $i = 1,2$, so V_1 and V_2 are orthogonal and $V = V_1 + V_2$. Clearly $\mathcal{R}_i$ is a root system in V_i. Now repeat as often as necessary.

Proposition 14.3 For any root system $\mathcal{R}$ in V and any $\alpha \in \mathcal{R}$, we have $(S_\alpha(x), S_\alpha(y)) = (x, y)$ for all $x, y \in V$.

PROOF From the definition we have

$$(S_\alpha(x), S_\alpha(y)) = \left(x - 2\frac{(\alpha, x)}{(\alpha, \alpha)}\alpha,\ y - 2\frac{(\alpha, y)}{(\alpha, \alpha)}\alpha\right)$$

$$= (x, y) - 4\frac{(\alpha, x)(\alpha, y)}{(\alpha, \alpha)} + 4\frac{(\alpha, x)(\alpha, y)(\alpha, \alpha)}{(\alpha, \alpha)(\alpha, \alpha)} = (x, y).$$

Proposition 14.4 For any root system $\mathscr{R}$ in V and $\alpha, \beta \in \mathscr{R}$, we have $2(\alpha, \beta)/(\beta, \beta) = 0, \pm 1, \pm 2$, or ± 3. Define $N(\alpha, \beta) = 2(\alpha, \beta)/(\beta, \beta)$.

PROOF By extending the base field of V to the real numbers R, we can consider $\mathscr{R}$ to be a subset of R^m, where m is the dimension of V and (x, y) on V is just the restriction of the usual inner product on R^m. If we let θ denote the angle between α and β thought of as vectors in R^m, we have

$$\cos^2\theta = (\alpha, \beta)^2/(\alpha, \alpha)(\beta, \beta)$$

so that $N(\alpha, \beta)N(\beta, \alpha) = 4\cos^2\theta \leq 4$. Since $N(\alpha, \beta)$ and $N(\beta, \alpha)$ are both integers, $-4 \leq N(\alpha, \beta) \leq 4$. We must rule out the possibility that $N(\alpha, \beta) = \pm 4$. Assume $N(\alpha, \beta) = 4$. Then $N(\beta, \alpha) = 1$ and $\cos\theta = 1$. Thus $(\alpha, \alpha) = 2(\alpha, \beta) = 4(\beta, \beta)$. However, then $\theta = 0$ and $(\alpha - 2\beta, \alpha - 2\beta) = 0$ so that $\alpha = 2\beta$, a contradiction. Similarly if $N(\alpha, \beta) = -4$, then $\theta = \pi$ and $\alpha = -2\beta$, again a contradiction.

Examples (2) The following is a list of the possible root systems when V is two-dimensional (see Fig. 14.1). In each case we describe a $V \subset R^2$, draw a graph, and assign to the root system a certain "type." The notation for the types follow a scheme that will be explained later in this section; also see the work of Samelson, [1969, p.47].

(i) Type $A_1 \times A_1$, $V = Q^2$, $\alpha = (1, 0)$, $\beta = (0, 1)$, and $\mathscr{R} = \{\pm\alpha, \pm\beta\}$.
(ii) Type A_2, $V = \{(s, 3^{1/2}t) : s, t \in Q\}$, $\alpha = (1, 0)$, $\beta = (-1/2, 3^{1/2}/2)$, and $\mathscr{R} = \{\pm\alpha, \pm\beta, \pm(\alpha + \beta)\}$.
(iii) Type B_2, $V = Q^2$, $\alpha = (1, 0)$, $\beta = (-1, 1)$, and $\mathscr{R} = \{\pm\alpha, \pm\beta, \pm(\alpha + \beta), \pm(2\alpha + \beta)\}$.
(iv) Type G_2, $V = \{(s, 3^{1/2}t : s, t \in Q\}$, $\alpha = (1, 0)$, $\beta = (-3/2, 3^{1/2}/2)$, and $\mathscr{R} = \{\pm\alpha, \pm\beta, \pm(\alpha + \beta), \pm(2\alpha + \beta), \pm(3\alpha + \beta), \pm(3\alpha + 2\beta)\}$.

Notice that the root system of type $A_1 \times A_1$ is not irreducible but the other three root systems are.

(3) Suppose we wish to compute the Weyl group of the above root system of type G_2. The short (or long) roots are the vertices of a hexagon. It is clear that the Weyl group W must be a subgroup of the dihedral group

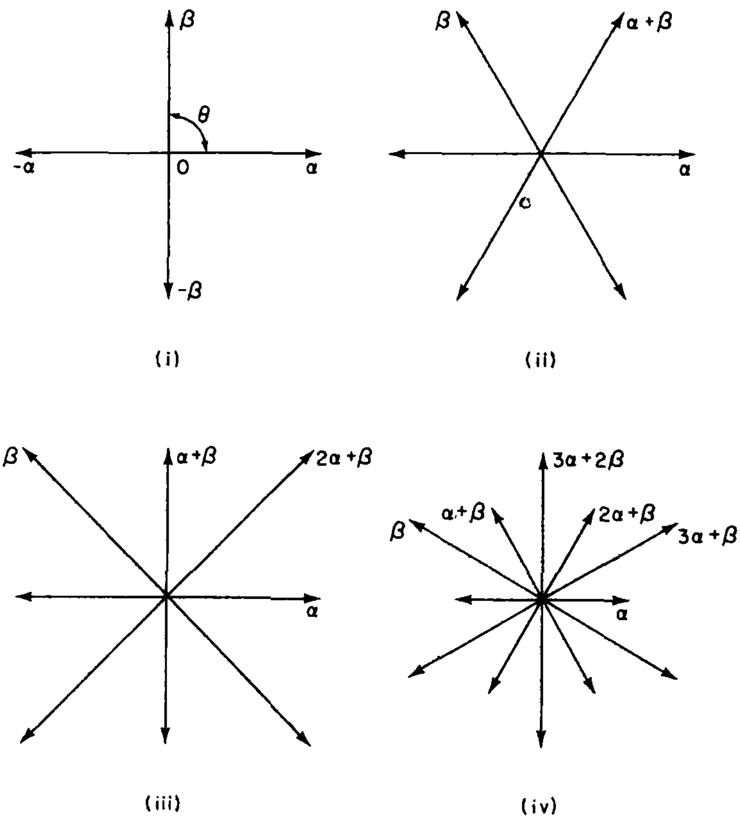


Fig. 14.1. Root systems. (i) Type $A_1 \times A_1$. (ii) Type A_2. (iii) Type B_2. (iv) Type G_2.

of order 12 consisting of all rotations and reflections of this hexagon. We see $S_\beta S_{\alpha+\beta}(\alpha) = S_\beta(2\alpha + \beta) = 2\alpha + \beta$ and $S_\beta S_{\alpha+\beta}(2\alpha + \beta) = S_\beta(\alpha) = \alpha + \beta$, so $S_\beta S_{\alpha+\beta}$ rotates the hexagon counterclockwise 60 degrees. Thus W is clearly the entire dihedral group of order 12.

Exercises (1) Find the Weyl groups of the other three root systems in example (2) above.

(2) Show that there are seven possibilities for the angle θ between roots α and $\beta \neq \pm\alpha$ of a root system, namely $\theta = \pi/6$, $\pi/4$, $\pi/3$, $\pi/2$, $2\pi/3$, $3\pi/4$, $5\pi/6$. Show that all of these angles occur in the graphs of the root systems of example (2). Show that if $N(\alpha, \beta) = \pm 2$, then $N(\beta, \alpha) = \pm 1$, $\theta = \pi/4$ or $3\pi/4$, and $(\alpha, \alpha) = 2(\beta, \beta)$. Also if $N(\alpha, \beta) = \pm 3$, then $N(\beta, \alpha) = \pm 1$, $\theta = \pi/6$ or $5\pi/6$, and $(\alpha, \alpha) = 3(\beta, \beta)$. Finally if $N(\alpha, \beta) = N(\beta, \alpha) = \pm 1$, then $\theta = \pi/3$ or $2\pi/3$, and $(\alpha, \alpha) = (\beta, \beta)$.

Proposition 14.5 If α and β are two roots in a root system $\mathcal{R}$ with $\beta \neq \pm\alpha$ and $(\alpha, \beta) > 0$, then $\alpha - \beta \in \mathcal{R}$.

PROOF Since $(\alpha, \beta) > 0$ implies $N(\alpha, \beta) > 0$ and $N(\beta, \alpha) > 0$, exercise (2) implies $N(\alpha, \beta) = 1$ or $N(\beta, \alpha) = 1$. If $N(\alpha, \beta) = 1$, then $S_\beta(\alpha) = \alpha - N(\alpha, \beta)\beta = \alpha - \beta \in \mathcal{R}$. If $N(\beta, \alpha) = 1$, then $-S_\alpha(\beta) = -\beta + N(\beta, \alpha)\alpha = \alpha - \beta \in \mathcal{R}$, recalling from remark (2) that $\alpha \in \mathcal{R}$ implies $-\alpha \in \mathcal{R}$.

Proposition 14.6 If $\alpha, \beta \in \mathcal{R}$, $\beta \neq \pm\alpha$, and p and q are the largest integers such that $\beta + p\alpha \in \mathcal{R}$ and $\beta - q\alpha \in \mathcal{R}$, then

(a) $\beta + k\alpha \in \mathcal{R}$ for all integers k with $-q \leq k \leq p$;
(b) $N(\beta, \alpha) = 2(\alpha, \beta)/(\alpha, \alpha) = q - p$.

PROOF (a) The result is obvious unless $p > 1$ or $q > 1$. Assume that $p > 1$. Then $-3 \leq N(\beta, \alpha) = 2(\alpha, \beta)/(\alpha, \alpha) \leq 3$ so that $-\frac{3}{2}(\alpha, \alpha) \leq (\alpha, \beta) \leq \frac{3}{2}(\alpha, \alpha)$, and for $k \geq 2$ we have $(\alpha, \beta + k\alpha) = (\alpha, \beta) + k(\alpha, \alpha) > 0$. By Proposition 14.5, $(\beta + k\alpha) - \alpha = \beta + (k-1)\alpha \in \mathcal{R}$ if $\beta + k\alpha \in \mathcal{R}$. Thus β, $\beta + \alpha$, $\beta + 2\alpha, \ldots, \beta + p\alpha \in \mathcal{R}$. By considering $-\alpha$ instead of α, we find $\beta - q\alpha$, $\beta - (q-1)\alpha, \ldots, \beta \in \mathcal{R}$.

(b) Since

$$S_\alpha(\beta + k\alpha) = \beta + k\alpha - 2[(\beta + k\alpha, \alpha)/(\alpha, \alpha)]\alpha = \beta - (k + N(\beta, \alpha))\alpha,$$

we must have $S_\alpha(\beta + p\alpha) = \beta - q\alpha$ and $\beta - (p + N(\beta, \alpha))\alpha = \beta - q\alpha$. Thus $N(\beta, \alpha) = q - p$ as required. Notice that this is precisely the result in Proposition 13.17 proved for root systems of split semisimple Lie algebras.

Definition 14.7 (a) A subset $\mathcal{B} \subset \mathcal{R}$ is called a **root system basis** for the root system $\mathcal{R}$ in V if $\mathcal{B}$ is a vector space basis for V and, for any $\beta \in \mathcal{R}$ we have $\beta = \sum_{i=1}^n m_i \alpha_i$ where $\mathcal{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, and either all the m_i's are nonnegative or they are all nonpositive.

(b) A root system basis $\mathcal{B} \subset \mathcal{R}$ is said to be **irreducible** if there is no nontrivial disjoint union $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ with $(\alpha, \beta) = 0$ for all $\alpha \in \mathcal{B}_1$ and $\beta \in \mathcal{B}_2$.

Proposition 14.8 Every root system possesses a root system basis.

PROOF Given the root system $\mathcal{R}$ in V, choose any vector space basis for V from $\mathcal{R}$, say $\{\beta_1, \beta_2, \ldots, \beta_n\}$. Any $\beta \in \mathcal{R}$ can be written uniquely as $\beta = \sum_{i=1}^n t_i \beta_i$ with $t_i \in Q$. We define a total ordering of $\mathcal{R}$ by prescribing $x > 0$ for $x \in V$ if the first nonzero coefficient of $x = \sum_{i=1}^n t_i \beta_i$ is positive, and for $\alpha, \beta \in \mathcal{R}$ set $\alpha > \beta$ if $\alpha - \beta > 0$. The usual properties for inequalities now hold. Define $\mathscr{R}^+ = \{\alpha \in \mathscr{R} : \alpha > 0\}$ and $\mathscr{R}^- = \{-\alpha : \in \mathscr{R}^+\} = \{\alpha \in \mathscr{R} : -\alpha > 0\}$. Finally define $\mathscr{B} = \{\alpha \in \mathscr{R}^+ : \text{for all } \beta, \gamma \in \mathscr{R}^+, \alpha \neq \beta + \gamma\}$. The properties of the next lemma show that $\mathscr{B}$ is a root system basis for $\mathscr{R}$.

Lemma 14.9 With the notation as above, we have the following.

(a) $\mathscr{B}$ spans V.
(b) If $\alpha, \beta \in \mathscr{B}$ and $\alpha \neq \beta$, then $(\alpha, \beta) \leq 0$.
(c) $\mathscr{B}$ is a vector space basis of V.
(d) If $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ and $\beta \in \mathscr{R}^+$, then either $\beta \in \mathscr{B}$ or there is some $\alpha_i \in \mathscr{B}$ with $\beta - \alpha_i \in \mathscr{R}^+$.
(e) If $\beta \in \mathscr{R}^+$, then there exist positive integers m_i for $i = 1, \ldots, n$ with $\beta = \sum_{i=1}^n m_i \alpha_i$.

PROOF (a) It suffices to show that $\mathscr{B}$ spans $\mathscr{R}^+$. Suppose $\alpha \in \mathscr{R}^+$ but is not in the span of $\mathscr{B}$, Then $\alpha = \beta + \gamma$ where $\beta, \gamma \in \mathscr{R}^+$. Now $\beta > \alpha, \gamma > \alpha$, and either β or γ is not in the span of $\mathscr{B}$. Thus by repeating this process we find there is no smallest element of $\mathscr{R}^+$ not in the span of $\mathscr{B}$, a contradiction.

(b) If $\alpha, \beta \in \mathscr{B}$, $\alpha \neq \beta$, and $(\alpha, \beta) > 0$, then by Proposition 14.5 either $\alpha - \beta \in \mathscr{R}^+$ or $\beta - \alpha \in \mathscr{R}^+$. However, then either $\alpha = (\alpha - \beta) + \beta$ with β, $\alpha - \beta \in \mathscr{R}^+$, or $\beta = (\beta - \alpha) + \alpha$ with $\alpha, \beta - \alpha \in \mathscr{R}^+$, a contradiction.

(c) We must show that the vectors of $\mathscr{B}$ are linearly independent. Suppose they are not. Then we can write $\sum_{i=1}^n t_i \alpha_i = 0$ where each $\alpha_i \in \mathscr{B}$, and the α_i's are ordered so that $t_i > 0$ for each $i \leq k$, $t_i \leq 0$ for $i > k$ and some $k \geq 1$. Then

$$0 < \left(\sum_{i=1}^k t_i \alpha_i, \sum_{i=1}^k t_i \alpha_i\right) = \left(\sum_{i=1}^k t_i \alpha_i, \sum_{j=k+1}^n (-t_j)\alpha_j\right)$$
$$= \sum_{i=1}^k \sum_{j=k+1}^n (-t_i t_j)(\alpha_i, \alpha_j) \leq 0$$

which is a contradiction.

(d) Assume β is a root such that $\beta \in \mathscr{R}^+$, $\beta \notin \mathscr{B}$, and $\beta - \alpha_i \notin \mathscr{R}^+$ for each $\alpha_i \in \mathscr{B}$. If for some α_i we have $\beta - \alpha_i \in \mathscr{R}^-$, then $\alpha_i = (\alpha_i - \beta) + \beta$ with $(\alpha_i - \beta)$, $\beta \in \mathscr{R}^+$, a contradiction. Thus $\beta - \alpha_i \notin \mathscr{R}$ for all $\alpha_i \in \mathscr{B}$, and by Proposition 14.5, $(\beta, \alpha_i) \leq 0$. Using this fact and (b), we can follow the proof of part (c) to show that $\{\beta, \alpha_1, \ldots, \alpha_n\}$ is linearly independent, a contradiction.

(e) This follows easily by induction using (d). Thus if $\beta \notin \mathscr{B}$, then there exists $\alpha_i \in \mathscr{B}$ with $\beta - \alpha_i \in \mathscr{R}^+$. Since $\beta - \alpha_i < \beta$, we can assume by induction that $\beta - \alpha_i = \sum n_k \alpha_k$ with n_k positive integers; this gives the results.

Proposition 14.10 If $\mathscr{B}$ is a root system basis for a root system $\mathscr{R}$ in V, then $\mathscr{B}$ is irreducible if and only if $\mathscr{R}$ is irreducible.

PROOF Suppose $\mathscr{B}$ is not irreducible so $\mathscr{B} = \mathscr{B}_1 \cup \mathscr{B}_2$ with $(\alpha, \beta) = 0$ for $\alpha \in \mathscr{B}_1$ and $\beta \in \mathscr{B}_2$. Let $\mathscr{R}_i$ be the roots in $\mathscr{R}$ spanned by $\mathscr{B}_i$, $i = 1, 2$. Then we claim that $\mathscr{R} = \mathscr{R}_1 \cup \mathscr{R}_2$, for suppose $\alpha \in \mathscr{R}$ but $\alpha \notin \mathscr{R}_1$ and $\alpha \notin \mathscr{R}_2$. Then by Lemma 14.9(d) we can assume that $\alpha \in \mathscr{R}$ with $\alpha = \alpha_i + \beta$, $\alpha_i \in \mathscr{B}_1$, and β in the span of $\mathscr{B}_2$. Now $S_{\alpha_i}(\alpha) = S_{\alpha_i}(\alpha_i + \beta) = -\alpha_i + \beta \in \mathscr{R}$, and this contradicts the definition of a root system basis. Conversely, suppose $\mathscr{R} = \mathscr{R}_1 \cup \mathscr{R}_2$ is not irreducible; then $\mathscr{B} = \mathscr{B}_1 \cup \mathscr{B}_2$ where $\mathscr{B}_i = \mathscr{R}_i \cap \mathscr{B}$. This shows that $\mathscr{B}$ is not irreducible.

Example (4) For each of the root systems in example (2), the notation has been chosen so that $\mathscr{B} = \{\alpha, \beta\}$ forms a root system basis.

Definition 14.11 Given a basis $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ for a root system $\mathscr{R}$ in V, let $N(\alpha, \beta) = 2(\alpha, \beta)/(\beta, \beta)$ for any $\alpha, \beta \in \mathscr{R}$ as in Proposition 14.4.

(a) The matrix $(N(\alpha_i, \alpha_j))$ is called the **Cartan matrix** of the root system $\mathscr{R}$.

(b) If $\alpha = \sum_{i=1}^{n} m_i\alpha_i \in \mathscr{R}^+$, then $\sum_{i=1}^{n} m_i$ is called the **height of** α (with respect to $\mathscr{B}$).

(c) Two root systems $\mathscr{R}_i$ in V_i, $i = 1, 2$, are said to be **isomorphic** if and only if there exists a nonsingular linear transformation T from V_1 onto V_2 such that $T(\mathscr{R}_1) = \mathscr{R}_2$ and $\langle Tx, Ty\rangle = c(x, y)$ for all $x, y \in V_1$ where $0 < c \in Q$ and (x, y) is the bilinear form for V_1 and $\langle x, y\rangle$ is the form for V_2.

Example (5) The Cartan matrices of the root systems in example (2) are

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}.$$

Also notice that the last three root systems have unique roots of maximal height.

REMARK (3) Notice that every Cartan matrix has 2's down the diagonal and negative integers or 0's elsewhere. Also notice that an isomorphism of root systems takes a basis of the first onto a basis of the second in such a way as to preserve the Cartan matrix.

Proposition 14.12 (a) All the roots of a root system $\mathscr{R}$ can be determined from a basis $\mathscr{B}$ for $\mathscr{R}$ and the Cartan matrix for $\mathscr{R}$ with respect to $\mathscr{B}$.

(b) Two root systems with bases such that their Cartan matrices are identical are isomorphic root systems.

PROOF (a) We will proceed by induction on the height of roots to find all roots of $\mathscr{R}^+$. The roots of height 1 are just those in $\mathscr{B}$. Assume we know the roots of $\mathscr{R}^+$ of height k and we wish to find the roots of height $k + 1$. By (d) of Lemma 14.9, every root of this height is of the form $\alpha + \alpha_i$ with $\alpha \in \mathscr{R}^+$ of height k and $\alpha_i \in \mathscr{B}$. The Cartan matrix allows us to compute

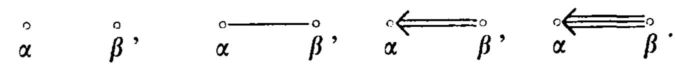
$N(\alpha, \alpha_i)$. By Proposition 14.6(b), $N(\alpha, \alpha_i) = q - p$ where $\alpha + k\alpha_i \in \mathscr{R}$ for $-q \le k \le p$. By the induction hypothesis q is known, so we are able to determine p and decide whether $\alpha + \alpha_i \in \mathscr{R}$ or not.

(b) This follows easily from (a).

Remark (4) At this stage we have not shown a given root system has the same Cartan matrix up to a permutation of rows and columns no matter what basis is chosen. This is in fact true but will only become clear once we show that essentially different Cartan matrices correspond to root systems of nonisomorphic Lie algebras.

Definition 14.13 The **Dynkin diagram** Δ of a root system $\mathscr{R}$ in V with basis $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ consists of a graph in the real space R^2 with n vertices labeled with $\alpha_1, \ldots, \alpha_n$ and $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i)$ line segments joining the vertex of α_i to the one α_j. Finally if $N(\alpha, \beta) \ne 0$ and $(\beta, \beta) > (\alpha, \alpha)$, draw an arrow on the line segments from the vertex of β to the vertex of α.

Example (6) The accompanying graphs are Dynkin diagrams of the four root systems of example (2).



Type $A_1 \times A_1$ Type A_2 Type B_2 Type G_2

Proposition 14.14 The Dynkin diagram Δ of a root system $\mathscr{R}$ in V with basis $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ completely determines the corresponding Cartan matrix of $\mathscr{R}$.

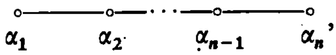
Proof We must merely show that $N(\alpha_i, \alpha_j)$ can be determined for $i \ne j$. If vertices α_i and α_j are not joined, then $N(\alpha_i, \alpha_j) = 0$. If they are joined by a single line, then $N(\alpha_i, \alpha_j) = N(\alpha_j, \alpha_i) = -1$ and $(\alpha_i, \alpha_i) = (\alpha_j, \alpha_j)$. If $N(\alpha_i, \alpha_j)N(\alpha_i, \alpha_j) = 2$, then the roots cannot have the same length with the longer being indicated by the arrow. Assume $(\alpha_j, \alpha_j) > (\alpha_i, \alpha_i)$. Then $-N(\alpha_i, \alpha_j) = -2(\alpha_i, \alpha_j)/(\alpha_j, \alpha_j) < -2(\alpha_i, \alpha_j)/(\alpha_i, \alpha_i) = -N(\alpha_j, \alpha_i)$, so $N(\alpha_i, \alpha_j) = -1$ and $N(\alpha_j, \alpha_i) = -2$. Finally if $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i) = 3$ and $(\alpha_j, \alpha_j) > (\alpha_i, \alpha_i)$, then a similar argument yields $N(\alpha_i, \alpha_j) = -1$ and $N(\alpha_j, \alpha_i) = -3$.

Remark (5) We have thus reduced the problem of finding all irreducible root systems to that of finding all Dynkin diagrams which (by Proposition 14.10) must have each vertex joined to at least one other vertex; that is, the diagram is connected.

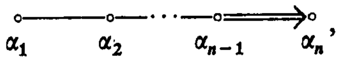
We shall show at the end of this section that there are only four infinite classes and five exceptional cases for Dynkin diagrams arising from irreducible root systems. These diagrams are listed in Proposition 14.15, and the notation used to label the diagram is historical with subscripts denoting the number of vertices and the letter the "shape" of the graph.

Proposition 14.15 The list of Dynkin diagrams arising from irreducible root systems $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ are precisely the accompanying.

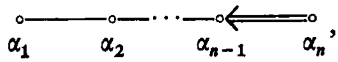
Type A_n, $n \geq 1$



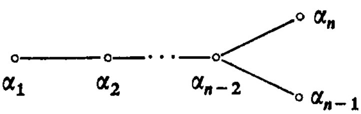
Type B_n, $n \geq 2$



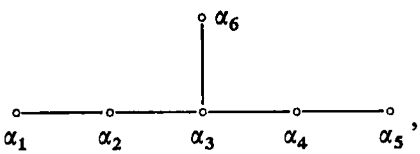
Type C_n, $n \geq 3$



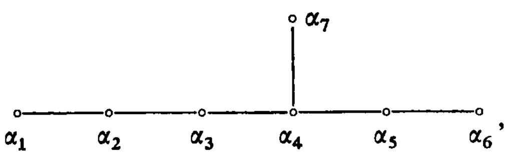
Type D_n, $n \geq 4$



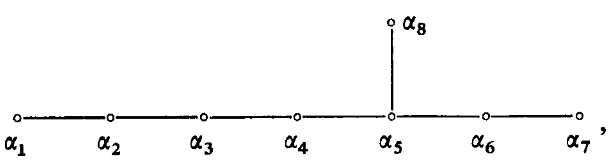
Type E_6



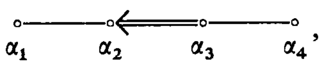
Type E_7



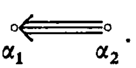
Type E_8



Type F_4



Type G_2



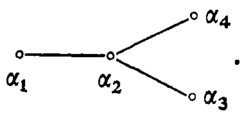
Example (7) Let $\alpha_1 = (-1, 0, 0, 0)$, $\alpha_2 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, $\alpha_3 = (0, -1, 0, 0)$, $\alpha_4 = (0, 0, -1, 0)$, and let $\mathscr{R}$ be the system of roots

$$\pm\alpha_i, \quad i = 1, 2, 3, 4; \qquad \pm(\alpha_2 + \alpha_i), \quad i = 1, 3, 4;$$
$$\pm(\alpha_1 + \alpha_2 + \alpha_3), \qquad \pm(\alpha_1 + \alpha_2 + \alpha_4), \qquad \pm(\alpha_2 + \alpha_3 + \alpha_4),$$
$$\pm(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4), \qquad \pm(\alpha_1 + 2\alpha_2 + \alpha_3 + \alpha_4).$$

Then $\mathscr{R}$ is an irreducible system of roots with $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ as a basis. The Cartan matrix and Dynkin diagram are

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & -1 \\ 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{bmatrix}$$

and



The Dynkin diagram is of type D_4.

We now prove Proposition 14.15 in many steps. First we shall find the diagrams without the arrows and put them in later; we shall still call these diagrams without arrows "Dynkin diagrams." Thus if $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ is a root system basis, we replace the α_i by the unit vector $X_i = \alpha_i/\|\alpha_i\|$ where $\|\alpha_i\|$ is the length of α_i determined by the form (x, y) on V considered as in R^n; note Definition 14.1. Consequently we study the set $\{X_1, \ldots, X_n\}$ satisfying

$$(X_i, X_i) = 1, \qquad (X_i, X_j) \le 0, \qquad 4(X_i, X_j)^2 = 0, 1, 2, 3 \qquad (*)$$

for $i \ne j$ and $i, j = 1, \ldots, n$. These conditions come from Lemma 14.9 and Proposition 14.4 using $\alpha_i = \|\alpha_i\| X_i$ so that $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i) = 4(X_i, X_j)^2$.

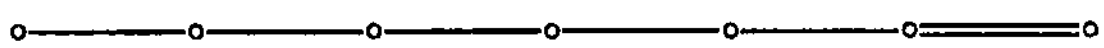
The corresponding Dynkin diagram, still denoted by Δ, consists of the points $X_1, \ldots, X_n$ in R^n as vertices with the number of lines joining X_i and X_j given by : X_i and X_j are connected by $4(X_i, X_j)^2 = 0, 1, 2$, or 3 line segments. Since we want to find the irreducible root systems, we see from remark (5) that the corresponding diagrams must be connected; that is, for points U, V in Δ there is a sequence $U_1 = U, U_2, \ldots, U_k = V$ in Δ so that U_i and U_{i+1} are connected in the diagram [note Proposition 14.6 and exercise (1), Section 14.2].

We now determine the connected diagrams using the following steps.

(1) Let Δ be a Dynkin diagram corresponding to the vectors $X_1, \ldots, X_n$. Let Δ' be the graph obtained by omitting a number of points and the lines

joining these points. Then Δ' is a Dynkin subdiagram corresponding to the remaining vectors $X_{i_1}, \ldots, X_{i_k}$.

This follows from the definition of a Dynkin diagram. For example,

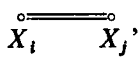


becomes

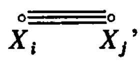
Using conditions (*) we also have the following.

(2) A correspondence between diagrams and values of inner products is given by

$(X_i, X_j) = -1/2,$



$(X_i, X_j) = -2^{1/2}/2,$



$(X_i, X_j) = -3^{1/2}/2.$

(3) There are no closed polygons.

PROOF Suppose $X_1, \ldots, X_k$ are the vertices of a polygon where X_i is connected to X_{i+1} for $1 \le i \le k-1$ and X_k is connected to X_1. Then set $X = \sum X_j$, and use (2) to compute

$$(X, X) = \sum (X_i, X_j) \le 0.$$

Thus $(X, X) = 0$ so that $\sum X_j = 0$. This contradicts the linear independence of $X_1, \ldots, X_n$ over Q.

(4) There are at most three lines coming from a vertex.

PROOF Let X be a vertex with $Y_1, \ldots, Y_k$ connected to X. Since there are no closed polygons, no two Y_i are connected. Thus $(Y_i, Y_j) = 0$ for $i \neq j$. In the vector space spanned by $X, Y_1, \ldots, Y_k$, we can choose a vector Y_0 so that $(Y_0, Y_0) = 1$ and $(Y_0, Y_i) = 0$ for $i = 1, \ldots, k$. We also have $(X, Y_0) \neq 0$, otherwise X is dependent on $Y_0, Y_1, \ldots, Y_k$ which implies X is dependent on $Y_1, \ldots, Y_k$, a contradiction to the choice of vertices in a diagram. Since $X = \sum (X, Y_i) Y_i$ is an orthogonal representation of the unit vector X, we have

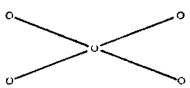
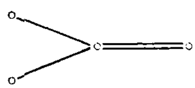
$$\begin{aligned} 1 = (X, X) &= (X, Y_0)^2 + (X, Y_1)^2 + \cdots + (X, Y_k)^2 \\ &> (X, Y_1)^2 + \cdots + (X, Y_k)^2. \end{aligned}$$

However, $4(X, Y_i)^2$ is the number of lines joining X and Y_i so that $4 > \sum_{i=1}^{k} 4(X, Y_i)^2$ which is the total number of lines coming from X.

(5) The only connected Dynkin diagram in which three lines join two vertices is of type G_2.

PROOF If this were not the case, there would be more than three lines coming from one vertex, contradicting (4).

We now consider the remaining cases where one or two lines join two vertices. Using (4), we note that the accompanying graphs are **not** Dynkin diagrams:

, , .

This can be be generalized as follows.

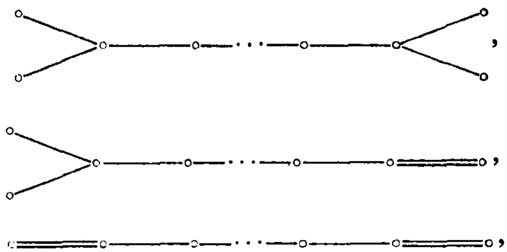
(6) Let $X_1, X_2, \ldots, X_i, X_{i+1}, \ldots, X_{i+k}, \ldots, X_n$ be a set of vectors satisfying condition (∗) and let Δ be the corresponding Dynkin diagram. Suppose the vertices $X_i, X_{i+1}, \ldots, X_{i+k}$ are such that X_{i+p} is connected to X_{i+p+1} by a single line for $p = 0, \ldots, k-1$ and let $X = \sum_{q=0}^{k} X_{i+q}$. Then the vectors $X_1, \ldots, X_{i-1}, X, X_{i+k+1}, \ldots, X_n$ satisfy condition (∗) and the corresponding Dynkin diagram Δ' is that of the original diagram except all the $X_i, \ldots, X_{i+k}$ have been replaced by the vector X.

PROOF Since X_{i+p} is connected to X_{i+p+1} by a single line, we have $2(X_{i+p}, X_{i+p+1}) = -1$. Since there are no closed polygons, $(X_{i+p}, X_{i+q}) = 0$ for $p < q$ unless $q = p + 1$. Thus we can compute

$$\begin{aligned}(X, X) &= \sum_{p,q=0}^{k} (X_{i+p}, X_{i+q}) \\ &= k + 2 \sum_{p<q} (X_{i+p}, X_{i+q}) \\ &= k + (k-1) = 1\end{aligned}$$

to obtain X as a unit vector. Now let Y be a vector in Δ with $Y \neq X_{i+p}$. Then since there are no closed polygons, Y is connected to at most one X_{i+r} for some r. Thus $(Y, X_{i+p}) = 0$ if $p \neq r$ which implies $(X, Y) = \sum (X_{i+p}, Y) = (X_{i+r}, Y)$. Therefore $4(X, Y)^2 = 4(X_{i+r}, Y)^2 = 0, 1, 2, 3$; that is, conditions (∗) are satisfied for $X_1, \ldots, X_{i-1}, X, X_{i+k+1}, \ldots, X_n$ and has Dynkin diagram as described.

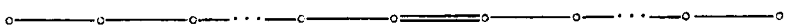
REMARK (6) In the following examples, the diagram Δ' is obtained from Δ by shrinking the vectors $X_i, \ldots, X_{i+k}$ (together with the single line joining them) to a single point X. If the accompanying graphs were Dynkin diagrams Δ, then they can be shrunk to Dynkin diagrams Δ' of the form following (5).

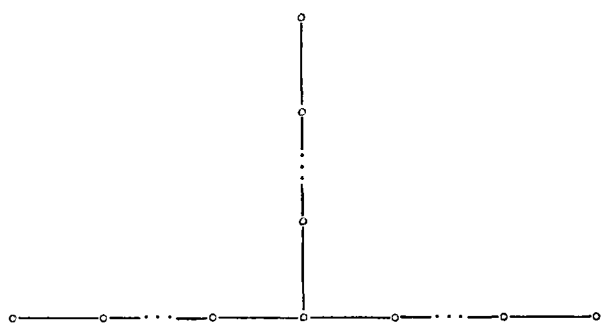


However, these small graphs are not Dynkin diagrams so the large graphs are not Dynkin diagrams.

(7) The only connected Dynkin diagrams are the accompanying ones.

 G_2,

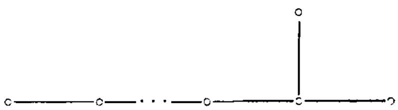




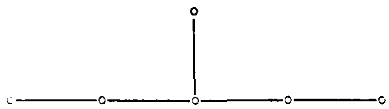
 A_n.

PROOF If a connected Dynkin diagram Δ has three lines joining two vertices, then Δ is of type G_2. If Δ has two lines joining two vertices, then, from remark (6), Δ is of the second type above. Finally we have the case where any two given vertices have one line joining them. Thus if there are two endpoints, we obtain a diagram of type A_n. If there are three endpoints, we obtain the third diagram. By remark (6) there cannot be four or more endpoints.

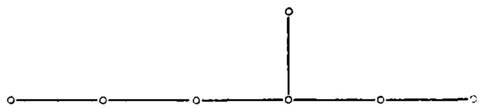
(8) If a connected Dynkin diagram has three endpoints, then it is one of the accompanying.



D_n ,



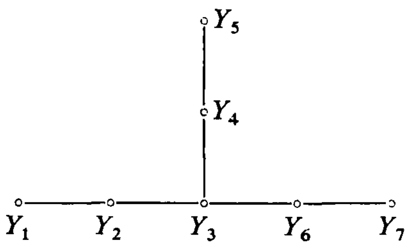
E_6 ,



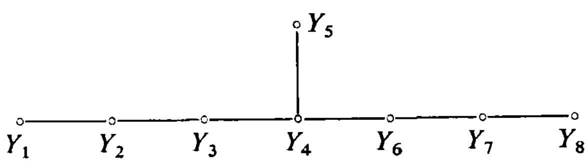
E_7 ,

E_8 ,

PROOF The third diagram in (7) is the only possibility, and one of the branches must contain only one vertex (besides the center vertex). If this is not the case, we use step (1) to obtain the accompanying subdiagram where



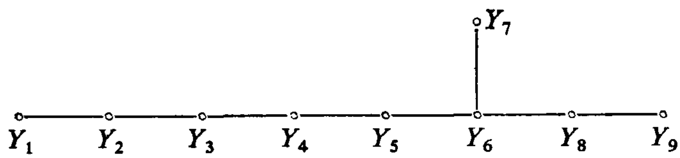
$\{Y_1, \ldots, Y_7\} \subset \{X_1, \ldots, X_n\}$. Let $X = Y_1 + 2Y_2 + 3Y_3 + 2Y_4 + Y_5 + 2Y_6 + Y_7$, and using step (2) we compute $(X, X) \leq 0$. Thus $X = 0$ which contradicts the linear independence of $Y_1, \ldots, Y_7$. Next a subdiagram of the form shown in the accompanying figure is impossible. Otherwise, let



$$X = Y_1 + 2Y_2 + 3Y_3 + 4Y_4 + 2Y_5 + 3Y_6 + 2Y_7 + Y_8$$

and we obtain $(X, X) \leq 0$ which contradicts the independence of $Y_1, \ldots, Y_8$. Thus only two vertices can occur at one end.

Finally the form of the accompanying subdiagram is impossible. Other-



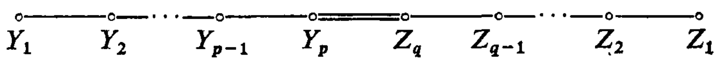
wise, $X = Y_1 + 2Y_2 + 3\,Y_3 + 4Y_4 + 5Y_5 + 6Y_6 + 3Y_7 + 4Y_8 + 2Y_9$ yields $(X, X) \leq 0$ and a contradiction. Thus the possibilities where one branch has one vertex (the endpoint) and the other branches have two or more vertices are of type E_6, E_7, E_8. When two branches have only one vertex (which must be an endpoint) we obtain a diagram of type D_n.

(9) The Dynkin diagrams which contain two vertices joined by two lines must be one of the accompanying.

B_n, C_n,

F_4

Proof From step (7) we must have the accompanying diagram where



the Y's and Z's are just a relabeling of the original X's. Let $Y = \sum_{i=1}^{p} iY_i$ and $Z = \sum_{k=1}^{q} kZ_k$. Then using $2(Y_i, Y_{i+1}) = -1 = 2(Z_k, Z_{k+1})$ and the fact other obvious vertices are not joined, we compute

$$\begin{aligned}(Y, Y) &= \sum_{i,j} (iY_i, jY_j) \\ &= \sum_{i=1}^{p} i^2 + 2 \sum_{i<j} ij(Y_i, Y_j) \\ &= \sum_{i=1}^{p} i^2 - \sum_{i=1}^{p-1} i(i+1) \\ &= p^2 - p(p-1)/2 = p(p+1)/2\end{aligned}$$

and

$$(Z, Z) = q(q+1)/2.$$

Next using certain vertices are not connected and step (2), we compute

$$\begin{aligned}(Y, Z) &= \sum_{i,k} ik(Y_i, Z_k) \\ &= pq(Y_p, Z_q) = -pq2^{1/2}/2.\end{aligned}$$

From Schwarz's inequality (and the independence of Y and Z),

$$p^2q^2/2 = (Y,Z)^2 < \|Y\|^2\|Z\|^2 = [p(p+1)/2]/[q(q+1)/2],$$

and since $pq > 0$ we obtain $2pq < (p+1)(q+1)$; that is, $(p-1)(q-1) < 2$. Because p and q are positive integers, the only possibilities are $p = 1$ with q arbitrary, $q = 1$ with p arbitrary, and $p = q = 2$. This gives the desired diagrams.

We now put the arrows on the diagrams replacing the unit vectors X_i by the $\alpha_i = \|\alpha_i\| X_i$ for $i = 1, \ldots, n$ in the root system basis $\mathscr{B}$. Since multiplying every α_i by a nonzero real number does not change the situation, we assume one of the α_i is a unit vector which we choose as an endpoint. Also note that α_i and α_j are connected by the same number of lines as X_i and X_j. Thus if $\alpha, \beta \in \mathscr{B}$ are connected by one line

$$N(\alpha, \beta)N(\beta, \alpha) = 1,$$

and since $N(\alpha, \beta)$ and $N(\beta, \alpha)$ are negative integers, we have

$$-1 = N(\alpha, \beta) = 2(\alpha, \beta)/(\beta, \beta) = N(\beta, \alpha) = 2(\beta, \alpha)/(\alpha, \alpha)$$

which implies $(\alpha, \alpha) = -2(\alpha, \beta) = (\beta, \beta)$. Thus there are no arrows on the Dynkin diagrams of type A_n, D_n, E_6, E_7, E_8.

If $\alpha, \beta \in \mathscr{B}$ are connected by two lines, then $N(\alpha, \beta)N(\beta, \alpha) = 2$ which gives two possibilities : $N(\alpha, \beta) = -1$, $N(\beta, \alpha) = -2$ or $N(\alpha, \beta) = -2$, $N(\beta, \alpha) = -1$ so that $(\beta, \beta) = 2(\alpha, \alpha)$ or $(\alpha, \alpha) = 2(\beta, \beta)$. Thus depending upon the choice of endpoint as a unit vector and using the results above for vertices joined by a single line, we obtain diagrams of type B_n, C_n, and F_4 (since F_4 is a symmetric graph, the choice of unit vector is immaterial). Similarly we obtain the diagram of type G_2 ; this completes the proof of Proposition 14.15.

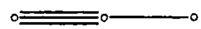
We have followed Jacobson [1962] and Samelson [1969] for the above proof and other variations are by Bourbaki [1968, Chap. 6] and Hausner and Schwartz [1968]. A method for choosing the vector X with $(X,X) \leq 0$ is given by Samelson [1969]. However with a little practice, it is easy to see how to make a guess for the desired coefficients; see exercise (6).

Exercises (3) Find a root system in R^3 of type A_3, that is, whose Dynkin diagram is of type A_3.

(4) If $\mathscr{R}$ is any root system and $\alpha \in \mathscr{R}$, set $\alpha' = 2\alpha/(\alpha, \alpha)$. Show that $S_{\alpha'} = S_\alpha$ and that $N(\alpha', \beta') = N(\beta, \alpha)$. Let $\mathscr{R}^* = \{\alpha' : \alpha \in \mathscr{R}\}$ and show that $\mathscr{R}^*$ is a root system called the **dual system** of $\mathscr{R}$. If $\mathscr{B}$ is a basis for $\mathscr{R}$, show that $\mathscr{B}^* = \{\alpha' : \alpha \in \mathscr{B}\}$ is a basis for $\mathscr{R}^*$. Finally show that if $\mathscr{R}$ is of type B_n, then $\mathscr{R}^*$ is of type C_n and conversely.

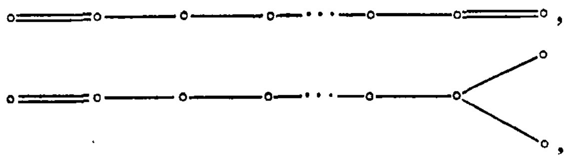
(5) Complete the determination of the arrows in the diagram of type G_2.

(6) (i) Show directly, without the above classification, that a diagram of type G_2 is not contained in any larger connected diagram. [Hint : Assume a diagram



with three vertices X_1, X_2, X_3. Then consider $X = rX_1 + sX_2 + tX_3$ and use step (2) to find r, s, t so that $(X, X) \leq 0$.]

(ii) As above show directly that the accompanying diagrams are impossible.



2. Classification of Split Simple Lie Algebras

In this section we will construct split simple Lie algebras corresponding to the Dynkin diagrams of type A_n, B_n, C_n, D_n, G_2, and F_4. Let g denote a split semisimple Lie algebra over a field K of characteristic 0, h a Cartan subalgebra, $\mathscr{R}$ the set of nonzero roots with Δ its Dynkin diagram, and we continue the previous notation.

Proposition 14.16 The algebra g is simple if and only if $\mathscr{R}$ is an irreducible root system.

PROOF Let $\alpha, \beta \in \mathscr{R}$. If $(\alpha, \beta) = \mathrm{Kill}(H_\alpha, H_\beta) = 0$ then $[H_\alpha Y] = \beta(H_\alpha)Y = (\alpha, \beta)Y = 0$ for $Y \in g(\beta)$. Suppose $\mathscr{R}$ is not irreducible so $\mathscr{R} = \mathscr{R}_1 \cup \mathscr{R}_2$ with $(\alpha, \beta) = 0$ if $\alpha \in \mathscr{R}_1$, $\beta \in \mathscr{R}_2$. Set h_i as the span of all H_α with $\alpha \in \mathscr{R}_i$ and $g_i = h_i + \sum_{\alpha \in \mathscr{R}_i} g(\alpha)$, $i = 1, 2$. From above it is clear that $[h_1 g_2] = 0$ and $[h_2 g_1] = 0$. Now consider $X \in g(\alpha)$, $Y \in g(\beta)$ with $\alpha \in \mathscr{R}_1$ and $\beta \in \mathscr{R}_2$. Since $(\alpha, \alpha + \beta) = (\alpha, \alpha) \neq 0$ and $(\beta, \alpha + \beta) = (\beta, \beta) \neq 0$, we have $\alpha + \beta \notin \mathscr{R}$ so $[XY] = 0$ and

$[g_1 g_2] = 0$. Thus $g = g_1 \oplus g_2$, a direct sum of ideals. The argument is reversible if g is not simple.

Exercise (1) Show g is simple if and only if the Dynkin diagram Δ (corresponding to $\mathscr{R}$ and $\mathscr{B}$) is connected.

Proposition 14.17 Let g be a split semisimple Lie algebra with nonzero root system $\mathscr{R}$, and let $\mathscr{B}$ be a root system basis for $\mathscr{R}$. For each $\alpha_i \in \mathscr{B}$, set

$$H_i = H'_{\alpha_i} = 2H_{\alpha_i}/(\alpha_i, \alpha_i).$$

Choose any $0 \neq X_i \in g(\alpha_i)$ and then let $Y_i \in g(-\alpha_i)$ be the unique element such that $[X_i Y_i] = H_i$. Now let

$$\mathscr{A} = \{H_i, X_i, Y_i : i = 1, 2, \ldots, n\},$$

where n is the dimension of h (and which equals the number of roots in $\mathscr{B}$). Then

(a) $\mathscr{A}$ generates g;

(b) $\mathscr{A}$ is a part of a basis for g such that all the basis elements not in $\mathscr{A}$ can be determined from Δ as well as all of the structure constants relative to this basis;

(c) The structure constants of the basis above are all rational numbers and any such basis for g is called a **Weyl basis**.

PROOF We will show all of the properties of the proposition simultaneously by showing by induction on the height of roots that we can step-by-step add basis vectors from higher root spaces which continue to satisfy the desired properties.

First a basis for root spaces of height 1 is contained in $\mathscr{A}$. We can compute explicitly

$$\begin{aligned}
[H_i H_j] &= 0, \\
[H_i X_j] &= N(\alpha_j, \alpha_i) X_j, \\
[H_i Y_j] &= -N(\alpha_j, \alpha_i) Y_j, \\
[X_i Y_i] &= H_i, \\
[X_i Y_j] &= 0 \qquad \text{if} \quad i \neq j.
\end{aligned}$$

Also $[X_i X_j]$ and $[Y_i Y_j]$ are in higher root spaces. Here the root space $g(-\alpha)$ will always be considered at the same time as $g(\alpha)$. Notice that the nontrivial structure constants are entries of the Cartan matrix and ca:. be determined from the Dynkin diagram Δ.

Next assume for all $\alpha \in \mathscr{R}^+$ with α of height k or less that $X_\alpha \in g(\alpha)$ and $Y_\alpha \in g(-\alpha)$ have been defined so that for any product of these elements which also lies in such a root space, all of the properties of the proposition hold. For any $\beta \in \mathscr{R}^+$ of height $k + 1$, we can write $\beta = \alpha_i + \alpha$ with $\alpha_i \in \mathscr{B}$

and α of height k using Lemma 14.9. Define $X_\beta = [X_i X_\alpha]$ and $Y_\beta = [Y_i Y_\alpha]$. Suppose in a similar manner $\gamma = \alpha_j + \delta$. We can now list some products which shows that the products at this height also satisfy the required properties. Thus using the Jacobi identity, we obtain

$$\begin{aligned}[X_\beta Y_\beta] &= [X_i[X_\alpha[Y_i Y_\alpha]]] + [[X_i[Y_i Y_\alpha]]X_\alpha] \\ &= [X_i[[X_\alpha Y_i]Y_\alpha]] + [X_i[Y_i[X_\alpha Y_\alpha]]] + N(\alpha, \alpha_i)[X_\alpha Y_\alpha] \\ &\quad + [[Y_i[X_i Y_\alpha]]X_\alpha].\end{aligned}$$

There is a similar formula for

$$\begin{aligned}[X_\gamma Y_\beta] &= [[X_j X_\delta][Y_i Y_\alpha]], \\ [X_i[X_j X_\alpha]] &= [X_j[X_i X_\alpha]] + [[X_i X_j]X_\alpha], \\ [H_i X_\gamma] &= N(\gamma, \alpha_i)X_\gamma.\end{aligned}$$

This completes the proof.

The following is an easy consequence of the preceding result:

Corollary 14.18 For any split simple Lie algebra g over a field K of characteristic 0, there exists a subset $g_0 \subset g$ such that g_0 is a split simple Lie algebra over Q. The root system of g_0 is the same as that of g, and $g \cong g_0(K) = g_0 \otimes_Q K$.

Theorem 14.19 Given a field K of characteristic 0, then:

(a) there exist split simple Lie algebras which have root systems determined by each of the Dynkin diagrams listed in Proposition 14.15;

(b) two such algebras are isomorphic if and only if they produce the same Dynkin diagram.

(c) the algebras are described in Table 14.1 where we use the notation $\mathscr{A}(n, K) = sl(n+1, K)$, $\mathscr{B}(n, K)$, $\mathscr{C}(n, K)$, $\mathscr{D}(n, K)$ for the algebras defined in Chapter 13.

We separate the discussion of Theorem 14.19 into several parts and give

TABLE 14.1

Root system	Model	Rank	Dimension
A_n	$\mathscr{A}(n, K)$	$n \geq 1$	$(n+2)n$
B_n	$\mathscr{B}(n, K)$	$n \geq 2$	$(2n+1)n$
C_n	$\mathscr{C}(n, K)$	$n \geq 3$	$(2n+1)n$
D_n	$\mathscr{D}(n, K)$	$n \geq 4$	$(2n-1)n$
E_6		6	78
E_7		7	133
E_8		8	248
F_4	$\mathscr{D}(M_3{}^8)$	4	52
G_2	$\mathscr{D}(\mathscr{C})$	2	14

some examples. First we assume part (a) and now consider (b). Let g_1 and g_2 be two split semisimple Lie algebras and suppose $\phi : g_1 \to g_2$ is an isomorphism. Then Cartan subalgebras, weight spaces, etc., of g_1 map onto a corresponding item in g_2. Also if $\mathrm{Kill}_i(\ ,\ )$ denotes the Killing form in g_i, then $\mathrm{Kill}_2(\phi X, \phi Y) = \mathrm{Kill}_1(X, Y)$. To see this we just note that from $\phi[XY] = [\phi X \phi Y]$ that $\operatorname{ad} \phi X = \phi \operatorname{ad} X \phi^{-1}$. Now take the trace according to the definition of the Killing form. The Dynkin diagrams are determined by the number of roots (which is the same for g_1 and g_2) and the numbers $N_i(\alpha, \beta)$ for $i = 1, 2$ using Proposition 14.14. These numbers are in turn determined by the Killing forms which are equal as above. Thus the Dynkin diagrams are the same up to possible relabeling.

Conversely, suppose g_1 and g_2 have the same Dynkin diagram. Then Proposition 14.17 applied to the corresponding sets of generators $\mathscr{A}_1$ and $\mathscr{A}_2$ show that we obtain the same structure constants for a basis determined by $\mathscr{A}_1$ and $\mathscr{A}_2$. Choosing such a basis, we see that the linear map which assigns the corresponding basis elements is an isomorphism; also note the proof of Samelson [1969, p. 52] for a similar argument.

Example (1) When part (a) is proved, we shall see that $g = sl(3, K)$ has a root system of type A_2. To illustrate the details of constructing g solely from the Dynkin diagram, we apply Proposition 14.17. First we have the accompanying Dynkin diagram and the Cartan matrix

$$\underset{\alpha_1}{\circ}\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\underset{\alpha_2}{\circ}$$

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

where $N(\alpha_1, \alpha_2) = N(\alpha_2, \alpha_1) = -1$. Also $N(\alpha_1, \alpha_2) = q - p$ where $\alpha_1 + k\alpha_2 \in \mathscr{R}$ for $-q \le k \le p$. However $q = 0$ so $p = 1$ and $\alpha_1 + \alpha_2 \in \mathscr{R}$. We can assume $(\alpha_1, \alpha_1) = (\alpha_2, \alpha_2) = 1$ so $(\alpha_1, \alpha_2) = -\frac{1}{2}$, and we can compute $(\alpha_1 + \alpha_2, \alpha_1 + \alpha_2) = 1$, $(\alpha_1 + \alpha_2, \alpha_1) = (\alpha_1 + \alpha_2, \alpha_2) = \frac{1}{2}$. This gives

$$N(\alpha_1 + \alpha_2, \alpha_1) = N(\alpha_1 + \alpha_2, \alpha_2) = 1.$$

Thus $q = 1$ and $p = 0$ in these cases, so $2\alpha_1 + \alpha_2 \notin \mathscr{R}$, $\alpha_1 + 2\alpha_2 \notin \mathscr{R}$. Let $\{X_1, H_1, Y_1, X_2, H_2, Y_2\}$ be as in the proof of Proposition 14.17. For $\alpha_1 + \alpha_2$, the root of height 2, set $X_3 = [X_1 X_2]$, $Y_3 = [Y_1 Y_2]$. We can compute

$$\begin{aligned}
[X_3 Y_3] &= [X_1[[X_2 Y_1]Y_2]] + [X_1[Y_1[X_2 Y_2]]] + N(\alpha_2, \alpha_1)[X_2 Y_2] \\
&\quad + [[Y_1[X_1 Y_2]]X_2] \\
&= [X_1[Y_1 H_2]] - H_2 \\
&= -H_1 - H_2, \\
[Y_1 X_3] &= [[Y_1 X_1]X_2] + [X_1[Y_1 X_2]] = X_2, \\
[X_1 Y_3] &= Y_2.
\end{aligned}$$

Now let g be the Lie algebra spanned by $X_1, X_2, X_3, H_1, H_2, Y_1, Y_2, Y_3$ with multiplication determined as in Proposition 14.17 and listed in the

	X_1	X_2	X_3	H_1	H_2	Y_1	Y_2	Y_3
X_1	0	X_3	0	$-2X_1$	X_1	H_1	0	Y_2
X_2		0	0	X_2	$-2X_2$	0	H_2	$-Y_1$
X_3			0	$-X_3$	$-X_3$	$-X_2$	X_1	$-H_1-H_2$
H_1				0	0	$-2Y_1$	Y_2	$-Y_3$
H_2					0	Y_1	$-2Y_2$	$-Y_3$
Y_1			(*)			0	Y_3	0
Y_2							0	0
Y_3								0

accompanying tabulation, where (*) is determined by anticommutivity. Now g is a split Lie algebra of type A_2.

REMARK (1) Once we know that the Lie algebras of Theorem 14.19 exist, Proposition 14.17 could be used to construct an algebra of the proper type from the Dynkin diagram. The above example shows that if the Dynkin diagram is very large, there will be many computations involved.

We now consider parts (a) and (c) by giving models of algebras corresponding to a given Dynkin diagram. We do this for types A, B, C, D, G_2, and F_4. Most of the computations have already been done in the examples in Section 13.2. Since the Dynkin diagrams will be connected, these Lie algebra models are actually simple algebras.

Type A_n Let $g = sl(n+1, K)$ with $n \geq 1$ as in example (1), Section 13.2. Recall that it was determined that the roots of g were the set of all $\alpha_{jk} : h \to K$ such that $\alpha_{jk}(\sum_{i=1}^{n+1} a_{ii} E_{ii}) = a_{jj} - a_{kk}$ where $\sum_{i=1}^{n+1} a_{ii} = 0$. More precisely

$$\mathscr{R} = \{\alpha_{jk} : j \neq k, 1 \leq j, k \leq n+1\}$$

with $\alpha_{ij} + \alpha_{jk} = \alpha_{ik}$ and $\alpha_{jk} = -\alpha_{kj}$. Set $\alpha_1 = \alpha_{1k}$, $\alpha_2 = \alpha_{23}, \ldots, \alpha_n = \alpha_{nn+1}$, and let $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. We claim that $\mathscr{B}$ determines a Dynkin diagram of type A_n. Thus if $i < j$, then $\alpha_{jk} = \alpha_{jj+1} + \alpha_{j+1j+2} + \cdots + \alpha_{k-1k}$, and since $\alpha_{jk} = -\alpha_{kj}$, it is clear that $\mathscr{B}$ is a root system basis for $\mathscr{R}$.

It was also shown that

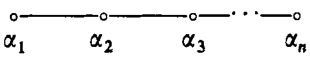
$$H_{\alpha_{jk}} = (E_{jj} - E_{kk})/2(n+1)$$

so

$$\begin{aligned}\langle \alpha_{ij}, \alpha_{km} \rangle &= \mathrm{Kill}((E_{ii} - E_{jj})/2(n+1), (E_{kk} - E_{mm})/2(n+1)) \\ &= [1/2(n+1)] \operatorname{trace} [(E_{ii} - E_{jj})(E_{kk} - E_{mm})].\end{aligned}$$

Thus by computing this trace, we see $\langle \alpha_{jk}, \alpha_{jk} \rangle = 1/(n+1)$ and $\langle \alpha_{ii+1}, \alpha_{i+1i+2} \rangle = -1/2(n+1)$. Finally $\langle \alpha_{ii+1}, \alpha_{jj+1} \rangle = 0$ for $j > i+1$.

To construct the Dynkin diagram, we join the n vertices $\alpha_1, \ldots, \alpha_n$ by various lines: α_i and α_j are joined by $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i)$ line segments where $N(\alpha, \beta) = 2\langle \alpha, \beta \rangle / \langle \beta, \beta \rangle$. From the above computations, we see the adjacent roots α_i and α_{i+1} are joined by $N(\alpha_i, \alpha_{i+1})N(\alpha_{i+1}, \alpha_i) = 1$ line segments, while nonadjacent roots α_i and α_j are joined by $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i) = 0$ line segments. If α_i and α_j are joined by a single line, then $N(\alpha_i, \alpha_j) = N(\alpha_j, \alpha_i) = -1$ and $\langle \alpha_i, \alpha_i \rangle = \langle \alpha_j, \alpha_j \rangle$. Thus there are no arrows. Consequently the Dynkin diagram for $sl(n+1, K)$ is



and so $sl(n+1, K)$ is a split simple Lie algebra of type A_n. A notation which would agree with the other notation in Chapter 13 is to define $\mathscr{A}(n, K) = sl(n+1, K)$.

To obtain the dimension given in Theorem 14.19, just use the dimension of $sl(n+1, K)$ as $(n+1)^2 - 1$. Also note Section 2.3 for arguments concerning the other dimensions.

Type B_n Let $\mathscr{B}(n, K)$ with $n \geq 2$ be the Lie algebra given in exercise (1), Section 13.2. Thus $\mathscr{B}(n, K)$ consists of all $(2n+1) \times (2n+1)$ matrices over K such that $J_n A^t J_n = -A$ where

$$J_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & O_n & & I_n \\ \vdots & & & \\ 0 & I_n & & O_n \end{bmatrix};$$

that is, $\mathscr{B}(n, K)$ consists of all skew-symmetric matrices relative to the nondegenerate symmetric bilinear form determined by J_n on a $2n+1$-dimensional vector space over K.

From the table of roots for $\mathscr{B}(n, K)$, let $\alpha_1 = \alpha_{1,2}, \alpha_2 = \alpha_{2,3}, \ldots, \alpha_{n-1} = \alpha_{n-1,n}$, and $\alpha_n = \varepsilon_n$, and let $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$. Then a straightforward computation shows $\mathscr{B}$ is a root system basis. For example, let $H = \sum a_{i+1,i+1}(E_{i+1,i+1} - E_{i+n+1,i+n+1})$ be in the Cartan subalgebra h. Then for $p < q$

$$\begin{aligned} \alpha_{p,q}(H) &= a_{p+1,p+1} - a_{q+1,q+1} \\ &= a_{p+1,p+1} - a_{p+2,p+2} + a_{p+2,p+2} - a_{p+3,p+3} + \cdots \\ &\quad + a_{q,q} - a_{q+1,q+1} \\ &= (\alpha_p + \cdots + \alpha_{q-1})(H). \end{aligned}$$

Similarly $\beta_{p,n} = \alpha_p + \cdots + \alpha_{n-1} + 2\alpha_n$, etc.

Instead of using direct computations as in the A_n-type, we shall use Proposition 14.6 to determine the B_n-type Dynkin diagram. First let $H \in h$ and $\alpha_1, \ldots, \alpha_n$ be as in the preceding paragraph. Then we compute the numbers

$$N(\alpha_i, \alpha_j)$$

by computing the corresponding numbers p and q according to Proposition 14.6. Thus consider α_i and α_{i+1} for $i = 1, \ldots, n-2$. We shall show

$$\alpha_{i+1} - \alpha_i \qquad \text{and} \qquad \alpha_{i+1} + 2\alpha_i$$

are not roots while

$$\alpha_i \qquad \text{and} \qquad \alpha_{i+1} + \alpha_i$$

are roots. Consequently $q = 0$, $p = 1$, and $N(\alpha_i, \alpha_{i+1}) = -1$. Thus for $H \in h$ as above we see

$$\begin{aligned}(\alpha_{i+1} - \alpha_i)(H) &= \alpha_{i+1,\,i+2}(H) - \alpha_{i,\,i+1}(H) \\ &= 2a_{i+2,\,i+2} - a_{i+3,\,i+3} - a_{i+1,\,i+1}.\end{aligned}$$

Comparing this value with the possible values of roots given in exercise (1), Section 13.2, we see $\alpha_{i+1} - \alpha_i$ is not a root. Similarly

$$\begin{aligned}(\alpha_{i+1} + 2\alpha_i)(H) &= (\alpha_{i+1,\,i+2} + 2\alpha_{i,\,i+1})(H) \\ &= 2a_{i+1,\,i+1} - a_{i+2,\,i+2} - a_{i+3,\,i+3}\end{aligned}$$

which also shows $\alpha_{i+1} + 2\alpha_i$ is not a root. Next $\alpha_{i+1} + \alpha_i = \alpha_{i,\,i+2}$ is a root because

$$\begin{aligned}(\alpha_{i+1} + \alpha_i)(H) &= \alpha_{i+1,\,i+2}(H) + \alpha_{i,\,i+1}(H) \\ &= a_{i+1,\,i+1} - a_{i+3,\,i+3} = \alpha_{i,\,i+2}(H).\end{aligned}$$

Thus for the root α_i and α_{i+1}, we see $N(\alpha_i, \alpha_{i+1}) = -1$ as desired. Similarly we can show $N(\alpha_{i+1}, \alpha_i) = -1$ so that in the Dynkin diagram the vertices α_i and α_{i+1} are joined by $N(\alpha_i, \alpha_{i+1})N(\alpha_{i+1}, \alpha_i) = 1$ line segment.

Next for α_i and α_j with $j \neq i + 1 \neq n$ and $j \neq n$, we see

$$(\alpha_i - \alpha_j)(H) = a_{i+1,\,i+1} - a_{i+2,\,i+2} - a_{j+1,\,j+1} + a_{j+2,\,j+2}$$

which shows $\alpha_i - \alpha_j$ is not a root. Similarly $\alpha_i + \alpha_j$ is not a root so that $p = q = 0$ and $N(\alpha_i, \alpha_j) = 0$. Thus in this case the α_i and α_j are joined by $N(\alpha_i, \alpha_j)N(\alpha_j, \alpha_i) = 0$ line segments.

Consider α_{n-1} and α_n. A computation shows $\alpha_{n-1} - \alpha_n$ is not a root, but

$$(\alpha_{n-1} + 2\alpha_n)(H) = a_{n,\,n} + a_{n+1,\,n+1} = \beta_{n-1,\,n}(H),$$

so that $\alpha_{n+1} + 2\alpha_n = \beta_{n+1,\,n}$ is a root. Thus $q = 0$, $p = 2$, and $N(\alpha_{n-1}, \alpha_n) = -2$. A similar computation shows $N(\alpha_n, \alpha_{n-1}) = -1$ so that α_{n-1} and α_n are joined by $N(\alpha_{n-1}, \alpha_n)N(\alpha_n, \alpha_{n-1}) = 2$ line segments.

The roots $\alpha_1, \ldots, \alpha_{n-1}$ are all the same length, since

$$N(\alpha_i, \alpha_{i+1}) = N(\alpha_{i+1}, \alpha_i) = -1$$

implies [using Proposition 14.6(b)]

$$(\alpha_{i+1}, \alpha_{i+1}) = -2(\alpha_i, \alpha_{i+1}) = (\alpha_i, \alpha_i).$$

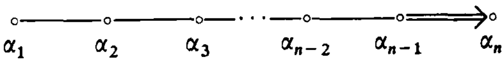
Next

$$-1 = N(\alpha_n, \alpha_{n-1}) = 2(\alpha_{n-1}, \alpha_n)/(\alpha_{n-1}, \alpha_{n-1})$$

and

$$-2 = N(\alpha_{n-1}, \alpha_n) = 2(\alpha_n, \alpha_{n-1})/(\alpha_n, \alpha_n)$$

implies $(\alpha_{n-1}, \alpha_{n-1}) = 2(\alpha_n, \alpha_n)$ so that α_{n-1} is longer than α_n. This gives an arrow from α_{n-1} to α_n and the accompanying Dynkin diagram.

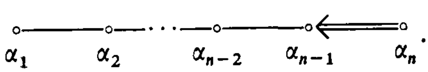


The same process can be used for the C_n-type and D_n-type which we now sketch and leave the computations to the reader.

Type C_n Let $\mathscr{C}(n, K)$ with $n \geq 3$ be the Lie algebra of exercise (3), Section 13.2. From the table of roots, we set

$$\alpha_1 = \alpha_{1,2}, \alpha_2 = \alpha_{2,3}, \ldots, \alpha_{n-1} = \alpha_{n-1,n}, \alpha_n = 2\delta_n.$$

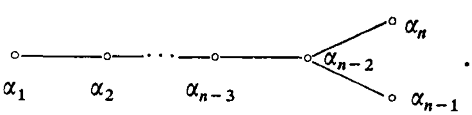
Then $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ is a root system basis. The roots $\alpha_1, \ldots, \alpha_{n-1}$ are the same length, but $(\alpha_n, \alpha_n) > (\alpha_{n-1}, \alpha_{n-1})$. Thus we obtain the accompanying Dynkin diagram.



Type D_n Let $\mathscr{D}(n, K)$ with $n \geq 4$ be the Lie algebra of exercise (2), Section 13.2. Then the root system of $\mathscr{D}(n, K)$ can be considered as consisting of precisely the roots $\alpha_{i,j}$, $\beta_{i,j}$, and $\gamma_{i,j}$ listed as roots of $\mathscr{B}(n, K)$. Thus if we let

$$\alpha_1 = \alpha_{1,2}, \ldots, \alpha_{n-1} = \alpha_{n-1,n}, \alpha_n = \beta_{n-1,n},$$

then $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ is a root system basis. The roots $\alpha_1, \ldots, \alpha_{n-2}$ are connected by a single line segment. Also the roots α_{n-2} and α_{n-1} are connected by a single line segment and so are the roots α_{n-2} and α_n. Thus we obtain the accompanying Dynkin diagram.



Exercises (2) Complete the computations to determine the Dynkin diagram for the C_n- and D_n-types.

(3) In the statement of Theorem 14.19, we make restrictions on the rank to obtain nonisomorphic algebras. When these restrictions are removed, show the following isomorphism between types; we denote these isomorphisms symbolically by

(i) $A_1 \cong B_1 \cong C_1$,
(ii) $B_2 \cong C_2$,
(iii) $A_3 \cong D_3$,
(iv) $D_2 \cong A_1 + A_1$; that is, a member of type D_2 is a direct sum of two of A_1 (see the book by Freudenthal and De Vries [1969] for details).

Next we consider some exceptional simple Lie algebras.

Type G_2 Consider the derivation algebra of a split Cayley algebra as described in example (2), Section 13.2. Thus let $g = \mathscr{D}(\mathscr{C}) = \{D(A, x, y) : A \in sl(3, K) \text{ and } x, y \in K^3\}$, and let $h = \{D(H, 0, 0) : H \in sl(3, K), H \text{ diagonal}\}$. It was shown that $\mathscr{R} = \{\pm \alpha_i : i = 1, 2, \ldots, 6\}$, where the α_i's can be described by setting $H = \sum_{i=1}^{3} a_{ii} E_{ii} \in sl(2, K)$ so $a_{11} + a_{22} + a_{33} = 0$ and listing the $\alpha_i(D(H, 0, 0))$ in the accompanying tabulation. Set $\alpha = \alpha_5$ and $\beta = \alpha_1$. Then $\mathscr{B} = \{\alpha, \beta\}$ is a root system basis for $\mathscr{R}$ since $\alpha + \beta = \alpha_4$, $2\alpha + \beta = -\alpha_6$, $3\alpha + \beta = \alpha_3$, $3\alpha + 2\beta = \alpha_2$, and $\mathscr{R} = \{\pm\alpha, \pm\beta, \pm(\alpha + \beta), \pm(2\alpha + \beta), \pm(3\alpha + \beta), \pm(3\alpha + 2\beta)\}$.

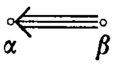
i	$\alpha_i(D(H, 0, 0))$
1	$a_{11} - a_{22}$
2	$a_{11} - a_{33}$
3	$a_{22} - a_{33}$
4	a_{11}
5	a_{22}
6	a_{33}

Now from this listing of roots, we see β, $\beta + \alpha$, $\beta + 2\alpha$, $\beta + 3\alpha$ are roots, but $\beta - \alpha$ is not a root. Therefore, from Proposition 14.6, we see $N(\beta, \alpha) = -3$. Similarly, since α and $\alpha + \beta$ are roots while $\alpha - \beta$ is not a root, we see $N(\alpha, \beta) = -1$. Thus α and β are joined by $N(\alpha, \beta)N(\beta, \alpha) = 3$ line segments. From

$$-3 = N(\beta, \alpha) = 2(\alpha, \beta)/(\alpha, \alpha)$$

we see $3(\alpha, \alpha) = -2(\alpha, \beta)$. Similarly $-1 = N(\alpha, \beta)$ implies $(\beta, \beta) = -2(\alpha, \beta)$.

Thus $(\beta, \beta) = 3(\alpha, \alpha)$ so that β is longer than α. This gives the accompanying Dynkin diagram so that $\mathscr{D}(\mathscr{C})$ is a split simple Lie algebra of type G_2.



From Section 9.6 we note that $\mathscr{C}$ is a simple nonassociative algebra with an identity element. Thus from exercise (3), Section 12.5, every derivation of $\mathscr{C}$ is inner. More explicitly, if P is the subspace of End($\mathscr{C}$) spanned by all $L(X)$ and $R(Y)$ for $X, Y \in \mathscr{C}$, then the Lie transformation algebra $\mathscr{L}(\mathscr{C}) = P + [P, P]$; see Section 7.2. Now the identities for $\mathscr{C}$ yield for any $X, Y \in \mathscr{C}$ that

$$D(X, Y) = [L(X), L(Y)] + [L(X), R(Y)] + [R(X), R(Y)]$$

is an inner derivation of $\mathscr{C}$, and consequently any derivation D of $\mathscr{C}$ equals $\sum_i D(X_i, Y_i)$ for suitable $X_i, Y_i \in \mathscr{C}$.

Type F_4 We again use derivations of a suitable nonassociative algebra to construct a Lie algebra of type F_4. First recall from example (3), Section 9.6 that the split Cayley algebra $\mathscr{C}$ given by 2×2 matrices has an involution $X \to \overline{X}$ given by

$$\begin{bmatrix} a & u \\ v & b \end{bmatrix} \to \begin{bmatrix} b & -u \\ -v & a \end{bmatrix}.$$

Now let $M_3{}^8$ denote the 3×3 Hermitian matrices with entries from $\mathscr{C}$ and with this involution. Thus $M_3{}^8$ is the set of all matrices

$$\begin{bmatrix} a_1 & X & \overline{Y} \\ \overline{X} & a_2 & Z \\ Y & \overline{Z} & a_3 \end{bmatrix}$$

where $a_i \in K$ and $X, Y, Z \in \mathscr{C}$. If $S, T \in M_3{}^8$, then the product

$$S \cdot T = \tfrac{1}{2}(ST + TS) \in M_3{}^8$$

where ST and TS is the usual matrix product. Thus $M_3{}^8$ is a **Jordan algebra**; that is, a nonassociative algebra which satisfies the identities

$$S \cdot T = T \cdot S \qquad \text{and} \qquad (S^2 \cdot T) \cdot S = S^2 \cdot (T \cdot S).$$

Let $J = M_3{}^8$ and let $g = \mathscr{D}(J)$. Then from the identities we see the Lie transformation algebra $\mathscr{L}(J) = P + [P, P]$ where P is the subspace spanned by all $L(X)$ for $X \in J$. Since J is a simple nonassociative algebra with an identity element 1, every derivation D of J is inner. Thus if $D = L(Z) + \sum [L(X_i), L(Y_i)]$, then $0 = D(1) = Z$ so that $D = \sum [L(X_i), L(Y_i)]$ for suitable $X_i, Y_i \in J$. We now sketch the proof that g is of type F_4 [Jacobson, 1971a, b, p. 407; Schafer, 1966].

Let E_{ij} denote the usual matrix basis and $E_i = E_{ii}$ for $i, j = 1, 2, 3$. Then from the above matrix description of J, we set

$$A_{ij} = AE_{ij} + \bar{A}E_{ji}$$

for $A \in \mathscr{C}$. Thus $A_{ij} \in J$ and

$$J_{ij} = \{A_{ij} : A \in \mathscr{C}\}$$

is a subspace such that $J_{ij} = J_{ji}$, and we have the direct sum

$$J = KE_1 + KE_2 + KE_3 + J_{12} + J_{13} + J_{23}.$$

Let $D_0 = \{D \in g : DE_i = 0 \text{ for } i = 1, 2, 3\}$. Then $D \in D_0$ implies $DJ_{ij} \subset J_{ij}$ using a few matrix computations involving $E_{ij} \cdot S$. Thus D induces an endomorphism of J_{ij}. This, in turn, induces an endomorphism D_{ij} of $\mathscr{C}$ by the formula $DA_{ij} = (D_{ij}\, A)_{ij}$ for $D \in D_0$, $A \in \mathscr{C}$, and $A_{ij} \in J_{ij}$ as above. From this we obtain the following result [Jacobson, 1971a; Schafer, 1966]. The Lie algebra L of all endomorphisms of the Cayley algebra $\mathscr{C}$ which are skew-symmetric relative to the bilinear form induced by the norm in $\mathscr{C}$ is of type D_4. Furthermore the map $D_0 \to L : D \to D_{ij}$ is an isomorphism of D_0 onto L.

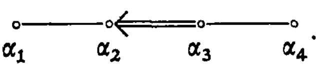
Using these results, we can decompose

$$g = D_0 + g_{12} + g_{13} + g_{23} \tag{$*$}$$

where $g_{12} = \{[L(E_1), L(A_{12})] : A \in \mathscr{C}\}$, $g_{13} = \{[L(E_1), L(A_{13})] : A \in \mathscr{C}\}$, and $g_{23} = \{[L(E_2), L(A_{23})] : A \in \mathscr{C}\}$. Now choose a four-dimensional Cartan subalgebra h for D_0 as discussed in type D_n and use ε_k to express the roots to obtain for $i \neq j$ 24 roots $\alpha_{i,j} = \varepsilon_i - \varepsilon_j$, $\beta_{i,j} = \varepsilon_i + \varepsilon_j$, $\gamma_{i,j} = -\varepsilon_i - \varepsilon_j$ for $1 \leq i, j \leq 4$. Thus since the dimension of the g_{ij} is 8, we obtain the dimension of g is 52.

In the above decomposition $(*)$ for g, $\operatorname{ad}_g h$ acts diagonally on g_{12}, g_{13}, and g_{23} and for $i = 1, 2, 3, 4$ has roots $\pm\varepsilon_i$; $\pm\Lambda_i$ where $\Lambda_i = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4) - \varepsilon_i$; and $\pm M_i$ where $M_1 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4)$, $M_2 = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4)$, $M_3 = \frac{1}{2}(\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4)$, $M_4 = \frac{1}{2}(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4)$. From this h is a Cartan subalgebra for g.

Next let $\alpha_1 = \frac{1}{2}(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4)$, $\alpha_2 = \varepsilon_4$, $\alpha_3 = \varepsilon_3 - \varepsilon_4$, $\alpha_4 = \varepsilon_2 - \varepsilon_3$. Then $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ is a root system basis. To obtain the Dynkin diagram, we use the usual process to compute the $N(\alpha_i, \alpha_j)$ noting we can add α_2 twice to α_3 and still have a root; this gives the accompanying diagram.



Exercise (4) Let A denote either of the algebras $\mathscr{C}$ or $M_3{}^8$ given by the above matrices, and let A_0 denote the subspace of matrices of trace 0. Show that $g = \mathscr{D}(A)$ acts irreducibly on A_0.

For an abstract detailed construction of algebras of type E, we refer to Hausner and Schwartz [1968] and for construction relating derivations and nonassociative algebras, details can be found in Jacobson [1971a, b] and Schafer [1966].

3. On Automorphisms of Simple Complex Lie Algebras

In this section we restrict ourselves to simple Lie algebras over the complex numbers and sketch results on automorphisms of their algebras. The algebras must automatically be split, so each must be of one of the types $A_n, \ldots, G_2$, etc.

If a closer study of the relationship of Cartan subalgebras and the corresponding root systems to automorphisms of split simple Lie algebras is made, then we find that the Dynkin diagram is also related to automorphisms of the algebra. We state without proof an important theorem about this relationship; this theorem is part of Theorem 33.9 of Freudenthal and de Vries [1969] and another variation is given by Jacobson [1962, Chap. 9].

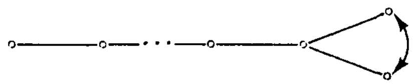
Theorem 14.20 Let g be a simple complex Lie algebra, let $\mathrm{Aut}(g)$ and $\mathrm{Int}(g)$ be the complex Lie groups of all automorphisms of g and the subgroup of inner automorphisms, let $\mathscr{B}$ be a root system basis for a system of roots $\mathscr{R}$ of g, and let Δ be the corresponding Dynkin diagram. Let $\mathrm{Aut}(\Delta)$ denote the group of automorphisms of Δ. More precisely $\mathrm{Aut}(\Delta)$ is the group of the permutations of $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ which preserve lengths of the α_i's and angles between pairs of them (this information is easily read from Δ). Then the group $\mathrm{Aut}(g)/\mathrm{Int}(g)$ is isomorphic to $\mathrm{Aut}(\Delta)$.

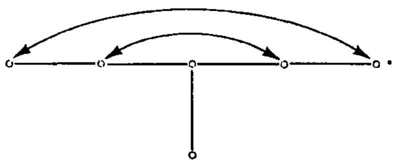
Proposition 14.21 Table 14.2 lists $\mathrm{Aut}(g)/\mathrm{Int}(g)$ for all simple Lie algebras g over the complex numbers.

TABLE 14.2

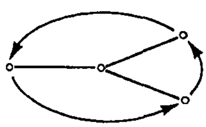
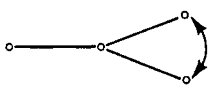
Type of g	$\mathrm{Aut}(g)/\mathrm{Int}(g)$
$A_n \quad n \geq 2$	Z_2
$B_n \quad n \geq 2$	$\{1\}$
$C_n \quad n \geq 3$	$\{1\}$
$D_n \quad n \geq 5$	Z_2
A_1, G_2, F_4, E_7, E_8	$\{1\}$
D_4	S_3
E_6	Z_2

PROOF The only Dynkin diagrams having one symmetry are those of types A_n, D_n, E_6. The nonidentity automorphism of Δ in these cases are given by arrows shown in the accompanying diagrams. The only Dynkin diagram

,
,

.

with more than one symmetry is D_4, and in this case the outside roots can be permuted in any manner. Thus two possibilities are illustrated in the accompanying diagrams.

,
.

Proposition 14.22 Let $g = sl(n, C)$ with $n \geq 2$ and $\varphi \in \mathrm{Aut}(g)$. Then there is some $U \in SL(n, C)$ such that

$$\varphi(X) = UXU^{-1} \qquad \text{for all } X \in g,$$

or

$$\varphi(X) = -UX^tU^{-1} \qquad \text{for all } X \in g.$$

PROOF It is easy to check that both definitions give automorphisms of g. If $Y \in g$, then $(\exp \operatorname{ad} Y)(X) = (\exp Y)X(\exp Y)^{-1}$ where $\exp Y \in SL(n, C)$. Thus if $\varphi \in \mathrm{Int}(g)$, then φ can be written as an automorphism of the first type above. Since, by Proposition 14.21, $\mathrm{Aut}(g)/\mathrm{Int}(g)$ has order 1 if $n = 2$ and has order 2 otherwise, we need only show that $X \to -X^t$ is not an inner automorphism if $n \geq 3$. Equivalently we need the fact that there is no $U \in SL\ (n, C)$ such that $UX = -X^tU$ for all $X \in sl(n, C)$. This is easy to show for $n \geq 3$ by choosing in succession the usual basis for g to substitute in for X and show that certain entries of U are 0 until one obtains a contradiction. For $n = 2$ we have the following result.

Exercise (1) Find a matrix $U \in SL(2, C)$ such that $UXU^{-1} = -X^t$ for all $X \in sl(2, C)$.

Proposition 14.23 Let $\mathscr{C}$ be a Cayley algebra over the complex numbers, let $g = \mathscr{D}(\mathscr{C})$ be its derivation algebra, and let $\mathrm{Aut}(\mathscr{C})$ be the automorphism group of $\mathscr{C}$. If $\varphi \in \mathrm{Aut}(g)$, then there exists some $T \in \mathrm{Aut}(\mathscr{C})$ with $\varphi(D) = TDT^{-1}$ for all $D \in g$.

PROOF Since g is of type G_2, we have $\mathrm{Aut}(g) = \mathrm{Int}(g)$. If $E \in \mathscr{D}(\mathscr{C})$, the exp $E \in \mathrm{Aut}(\mathscr{C})$ and $\exp(\mathrm{ad}\ E) \in \mathrm{Int}(g)$. However, $(\exp(\mathrm{ad}\, E))(D) = (\exp E)D(\exp E)^{-1}$, and since the maps $\exp(\mathrm{ad}\, E)$ generate $\mathrm{Int}(g)$, we obtain the results.

Exercises (2) F.ove a result analogous to Proposition 14.23 for the algebra $M_3{}^8$.

(3) Each of the simple Lie algebras $\mathscr{B}(n, C)$, $\mathscr{C}(n, C)$, $\mathscr{D}(n, C)$ can be described as a set of matrices X satisfying $JX^tJ^{-1} = -X$ for some fixed J. Consider the corresponding groups $B(n, C)$, $C(n, C)$, $D(n, C)$ as the set of nonsingular matrices U such that $JU^tJ^{-1} = U^{-1}$.

(i) Show that any inner automorphism of the three Lie algebras above can be written as $X \to UXU^{-1}$ with U in the corresponding group.

(ii) Show that for g of type B_n or C_n that $\mathrm{Aut}(g) = \mathrm{Int}(g)$; note Proposition 14.21.

(iii) For g of type D_n with $n \geq 4$, show $\mathrm{Aut}(g)$ has an automorphism which is not inner [Consider the map $X \to PXP^{-1}$ where P is given by the interchange map

$$(X_1, \ldots, X_{2n}) \to (X_1, \ldots, X_{n-1}, X_{n+1}, X_n, X_{n+2}, \ldots, X_{2n}).]$$

What can be said about $X \to -X^t$?

(iv) For $n > 4$ and g of type D_n, show $\mathrm{Aut}(g)/\mathrm{Int}(g)$ is of order 2. What is the general form of an element in $\mathrm{Aut}(g)$? Type D_4 is more complicated, and we refer to the work of Hausner and Schwartz [1968] and Jacobson [1971a].

CHAPTER 15

SIMPLE REAL LIE ALGEBRAS AND GROUPS

In Section 12.6 we showed that a finite-dimensional simple Lie algebra over the real numbers is either isomorphic to the realification of a simple complex Lie algebra or is isomorphic to a real form of a simple complex Lie algebra. We have classified the simple complex Lie algebra and have shown that a realification of a simple complex Lie algebra is a simple real Lie algebra. Thus it suffices to describe the real forms to obtain a classification of the simple Lie algebras over the real numbers, and we do this in Section 15.1. Next we consider the irreducible representations of simple real and complex Lie algebras discussing the maximal weight and how they are determined, the basic representations, and how an irreducible representation is obtained from these. We also discuss Weyl's formula for the dimension of an irreducible representation and give examples showing how the Dynkin diagram determines an irreducible representation. Finally in Section 15.3 we discuss how the previous results apply to Lie groups.

1. Real Forms of Simple Complex Lie Algebra

In this section we first discuss generalities about real forms of a simple complex Lie algebra $\tilde{g}$ by noting they are given in terms of a conjugation operator $C : \tilde{g} \to \tilde{g}$. These conjugations can be written in the form $C = \phi C_-$

where C_- is a conjugation of a compact real form and ϕ is a suitable automorphism of $\tilde{g}$ with $\phi^2 = I$. In the second part of this section, we discuss the real forms for the classical Lie algebras by explicitly computing the automorphisms ϕ. This discussion follows closely the general approach of Hausner and Schwartz [1968].

Definition 15.1 Let $\tilde{g}$ be a simple complex Lie algebra.

(a) If C is a conjugation of $\tilde{g}$, then the fixed point set $g = \{X \in \tilde{g} : C(X) = X\}$ is called the **real form of** C and C is called the **conjugation of** g.

(b) Two conjugations C_1 and C_2 of $\tilde{g}$ are called **equivalent conjugations** if the real forms of C_1 and C_2 are isomorphic.

Thus from results in Section 9.2 we have for the real form g the decomposition $\tilde{g} = g + ig$ and relative to this decomposition $C(X + iY) = X - iY$ for $X, Y \in g$. Also since the isomorphism of real forms defines an equivalence relation, it induces an equivalence relation on conjugations, and we seek a suitable representative for each equivalence class.

Proposition 15.2 Two conjugations C_1 and C_2 of a simple complex Lie algebra $\tilde{g}$ are equivalent if and only if there exists an automorphism φ of $\tilde{g}$ such that $C_2\varphi = \varphi C_1$.

Proof Let g_1 and g_2 be the real forms of C_1 and C_2. If φ is an automorphism of $\tilde{g}$ with $C_2\varphi = \varphi C_1$, then for all $X \in g_1$ we have $C_2(\varphi(X)) = \varphi(C_1(X)) = \varphi(X)$ so $\varphi(X) \in g_2$ and φ restricted to g_1 is an isomorphism onto g_2.

Conversely if C_1 and C_2 are equivalent so that g_1 and g_2 are isomorphic, choose a particular isomorphism $\psi : g_1 \to g_2$. Any element of $\tilde{g}$ can be written uniquely as $X + iY$ with $X, Y \in g_1$, so we can define $\varphi : \tilde{g} \to \tilde{g}$ by $\varphi(X + iY) = \psi(X) + i\psi(Y)$. For all $X, Y \in g_1$

$$\begin{aligned} C_2(\varphi(X + iY)) &= C_2(\psi(X) + i\psi(Y)) \\ &= \psi(X) - i\psi(Y) = \varphi(X - iY) = \varphi(C_1(X + iY)) \end{aligned}$$

so $C_2\varphi = \varphi C_1$ as required. It is easy to check that φ is an automorphism of $\tilde{g}$.

Proposition 15.3 Let $\tilde{g}$ be a simple complex Lie algebra.

(a) If C_1 and C_2 are conjugations of $\tilde{g}$, then C_1C_2 is an automorphism of $\tilde{g}$.

(b) If C_0 is any fixed conjugation of $\tilde{g}$ and C is any other conjugation, then $C = \varphi C_0$ where φ is an automorphism of $\tilde{g}$ such that $C_0\varphi C_0 = \varphi^{-1}$.

(c) If C_0 is a fixed conjugation of $\tilde{g}$ and $C_1 = \varphi_1 C_0$, $C_2 = \varphi_2 C_0$ are two conjugations as above, then C_1 and C_2 are equivalent if and only if there exists an automorphism φ of $\tilde{g}$ such that $\varphi_2 = \varphi\varphi_1 C_0 \varphi^{-1} C_0$.

PROOF (a) This consists of some straightforward computations which will be left to the reader.

(b) Since $C_0{}^2 = I$ we note $C = (CC_0)C_0$, so $C = \varphi C_0$ where $\varphi = CC_0$ is an automorphism of $\tilde{g}$. Now $I = C^2 = \varphi C_0 \varphi C_0$ so $C_0 \varphi C_0 = \varphi^{-1}$.

(c) By Proposition 15.2, C_1 and C_2 are equivalent if and only if there exists an automorphism φ with $\varphi_2 C_0 \varphi = C_2 \varphi = \varphi C_1 = \varphi\varphi_1 C_0$ or $\varphi_2 = \varphi\varphi_1 C_0 \varphi^{-1} C_0$.

REMARKS (1) Given a real form g of a simple complex Lie algebra $\tilde{g}$, the Killing form of g is equal to the Killing form of $\tilde{g}$ restricted to g, since a basis for g over the real numbers is also a basis for $\tilde{g}$ over the complex numbers This fact will be helpful in computing the Killing forms of real forms.

(2) The Killing form of a real form g is a nondegenerate symmetric bilinear form over the real numbers. Consequently by Sylvester's theorem, there exists an orthogonal basis for g such that $\mathrm{Kill}(X_i, X_i) = \pm 1$ for each basis element X_i, and the signature, or the number of $+1$'s minus the number of -1's, is independent of the basis chosen. The signature of the Killing form will be referred to as the **signature of** g. Notice that isomorphic real forms have equal signature; that is, more importantly, if two real forms have different signatures, then they are not isomorphic.

Definition 15.4 A real form g of a simple complex Lie algebra is called a **compact real form** if the Killing form of g is negative definite; that is, if the signature of g is $-m$ where m is the dimension of g.

Proposition 15.5 Any simple complex Lie algebra possesses at least two nonisomorphic real forms, a compact real form and a split real form. The signature of a split real form g is equal to the rank of g; that is, equal to the dimension of any Cartan subalgebra of g.

PROOF Given a simple complex Lie algebra $\tilde{g}$, choose a Cartan subalgebra $\tilde{h}$ of $\tilde{g}$ and a basis $\mathscr{B}$ for the root system $\mathscr{R}$ of $\tilde{g}$. By Proposition 14.17, $\tilde{g}$ possesses a Weyl basis; that is, a basis such that the Lie product of any two basis elements is equal to a linear combination of the basis elements with rational coefficients. Let g be the set of linear combinations of these basis elements with real coefficients. It is clear that g is a split real form of $\tilde{g}$; that is, a split real Lie algebra and a real form of $\tilde{g}$.

To compute the signature of g we construct an orthogonal basis for g

from the Weyl basis of g. Clearly one can choose an orthogonal basis for $h = g \cap \tilde{h}$, and the Killing form restricted to h is positive definite (note Proposition 13.16). Thus this part of the basis for g will contribute $+n$ to the signature of g where n equals the dimension of h. For each positive root $\alpha \in \mathscr{R}$ there exist $X_\alpha \in g(\alpha)$ and $Y_\alpha \in g(-\alpha)$ so that $[X_\alpha Y_\alpha] = H_\alpha' \in h$ with $[H_\alpha' X_\alpha] = 2X_\alpha$ and $[H_\alpha' Y_\alpha] = -2Y_\alpha$. However, then

$$\begin{aligned} a = 2\,\mathrm{Kill}(X_\alpha, Y_\alpha) &= \mathrm{Kill}([H_\alpha' X_\alpha], Y_\alpha) \\ &= \mathrm{Kill}(H_\alpha', [X_\alpha Y_\alpha]) = \mathrm{Kill}(H_\alpha', H_\alpha') > 0, \end{aligned}$$

so we can set

$$U_\alpha = (X_\alpha + Y_\alpha)/a^{1/2} \qquad \text{and} \qquad V_\alpha = (X_\alpha - Y_\alpha)/a^{1/2}.$$

Consequently $\mathrm{Kill}(U_\alpha, U_\alpha) = 2\,\mathrm{Kill}(X_\alpha, Y_\alpha)/a = 1$, $\mathrm{Kill}(V_\alpha, V_\alpha) = -1$, $\mathrm{Kill}(U_\alpha, U_\beta) = \mathrm{Kill}(U_\alpha, H) = \mathrm{Kill}(V_\alpha, H) = 0$ for all $H \in h$ and positive roots $\alpha, \beta \in \mathscr{R}$, and finally $\mathrm{Kill}(U_\alpha, U_\beta) = \mathrm{Kill}(V_\alpha, V_\beta) = 0$ for all positive roots $\alpha \neq \beta$. It is clear that the set of all U_α's and V_α's complete an orthogonal basis for g and contribute nothing to the signature of g. Thus the signature of g is n as required.

We can alter the above basis slightly to obtain a basis for a compact real form of $\tilde{g}$. Thus consider an orthogonal basis for $\tilde{g}$ as follows: iH_k for $k = 1, \ldots, n$ with $H_k \in h$ and iU_α, V_α as α varies over all positive roots of $\mathscr{R}$. Let g_1 denote the set of all linear combinations of these basis elements with real coefficients. Clearly the Killing form of $\tilde{g}$ restricted to g_1 is negative definite, so we must merely show that g_1 is a subalgebra of $\tilde{g}$ or, what is equivalent, that g_1 is the fixed point set of a conjugation of $\tilde{g}$. We see g_1 is the real form of the conjugation φC where C is the conjugation of the split real form g, and φ is an automorphism of $\tilde{g}$ such that $\varphi(H) = -H$ for all $H \in h$, and for any positive root α, $\varphi(X_\alpha) = -Y_\alpha$, $\varphi(Y_\alpha) = -X_\alpha$. It should be clear that X_α's and Y_α's can be chosen so that they satisfy the properties of the previous paragraph and also so that φ defines an automorphism of $\tilde{g}$; note Helgason [1962, p. 155] for a slightly different proof.

Example (1) Let $\tilde{g} = sl(n+1, C)$, so it is of type A_n. We have previously shown that, for $X, Y \in \tilde{g}$, $\mathrm{Kill}(X, Y) = 2(n+1)\,\mathrm{trace}(XY)$, and a Weyl basis is given by setting $H'_{k,k+1} = E_{k,k} - E_{k+1,k+1}$ for $k = 1, 2, \ldots, n$ and for $1 \le j < k \le n+1$, $X_{jk} = E_{jk}$, and $Y_{jk} = E_{kj}$. To find an orthogonal basis for the split real form, we first choose $H_1, H_2, \ldots, H_n \in h$ where h is the set of real diagonal matrices of trace 0 such that $\mathrm{Kill}(H_k, H_k) = 1$ and $\mathrm{Kill}(H_j, H_k) = 0$ if $j \neq k$. Also $U_{jk} = (E_{jk} + E_{kj})/2(n+1)^{1/2}$ and $V_{jk} = (E_{jk} - E_{kj})/2(n+1)^{1/2}$ for $1 \le j < k \le n+1$ and so clearly all these elements span the split real form $g = sl(n+1, R)$. The corresponding compact real form with basis iH_k, $k = 1, 2, \ldots, n$ and iU_{jk}, V_{jk}, $1 \le j < k \le n+1$ clearly is the algebra $su(n+1) = \{X \in \tilde{g} : X^{-t} = -X\}$.

Similarly in Table 15.1, $\tilde{g}$ is a simple classical Lie algebra over C of the type listed with n the rank of $\tilde{g}$ and g_- the compact form of $\tilde{g}$ up to isomorphism (note Proposition 15.6(b) and the book by Chevalley [1946] for a complete discussion of the compact $sp(n)$). The notation is given in Section 12.6, but we shall change it slightly later.

TABLE 15.1

$\tilde{g}$	g_-	dimension g_-
A_n $(n \geq 1)$	$su(n+1)$	$n(n+2)$
B_n $(n \geq 2)$	$so(2n+1)$	$n(2n+1)$
C_n $(n \geq 3)$	$sp(n)$	$n(2n+1)$
D_n $(n \geq 4)$	$so(2n)$	$n(2n-1)$

Proposition 15.6 Let $\tilde{g}$ be a simple complex Lie algebra, and let g_- be a compact real form with conjugation C_-.

(a) If C is any conjugation of $\tilde{g}$, then $C = \phi C_-$ where $\phi \in \mathrm{Aut}(\tilde{g})$, $\phi^2 = I$, and $C_- \phi = \phi C_-$.

(b) If K_- is a conjugation of $\tilde{g}$ corresponding to any other compact form of $\tilde{g}$, then C_- and K_- are equivalent.

PROOF (a) From Proposition 15.3, $C = \phi_1 C_-$ where $C_- \phi_1 C_- = \phi_1^{-1}$, so all we have to do is show there is a conjugation ϕC_- equivalent to C with $C_- \phi C_- = \phi$. Moreover we can assume that ϕ has the form $\phi = \psi\phi_1 C_- \psi^{-1} C_-$ for some $\psi \in \mathrm{Aut}(\tilde{g})$ which we give below.

For any $Z_1 = X_1 + iY_1$, $Z_2 = X_2 + iY_2 \in \tilde{g} = g_- + ig_-$ with X_i, $Y_i \in g_-$ we compute

$$\mathrm{Kill}(C_-(Z_1), C_-(Z_2)) = \mathrm{Kill}(X_1 - iY_1, X_2 - iY_2) = \overline{\mathrm{Kill}(Z_1, Z_2)} \qquad (*)$$

where $\bar{a}$ is the conjugate of a in the complex numbers. We define the form on $\tilde{g}$ by

$$(Z_1, Z_2) = -\mathrm{Kill}(Z_1, C_-(Z_2))$$

and see it is a positive definite Hermitian inner product on $\tilde{g}$ as follows. The form is linear in the first variable, additive in the second, and satisfies $-(Z_1, Z_2) = \mathrm{Kill}(Z_1, C_-(Z_2)) = \overline{\mathrm{Kill}(C_-(C_-(Z_1)), C_-(Z_2))} = \mathrm{Kill}(C_-(Z_2), C_-(C_-(Z_1))) = \overline{\mathrm{Kill}(Z_2, C_-(Z_1))} = -\overline{(Z_2, Z_1)}$, using $(*)$. Also $(Z_1, Z_1) = -\mathrm{Kill}(X_1 + iY_1, X_1 - iY_1) = -\mathrm{Kill}(X_1, X_1) - \mathrm{Kill}(Y_1, Y_1) \geq 0$ so that the form is positive definite.

For an endomorphism A of $\tilde{g}$ we define the adjoint A^* by $(AZ, W) = (Z, A^*W)$ and note that if $\mathrm{Kill}(AZ, AW) = \mathrm{Kill}(Z, W)$, then

$$\begin{aligned}(AZ, W) &= -\mathrm{Kill}(AZ, C_-(W)) \\ &= -\mathrm{Kill}(Z, A^{-1}C_-(W)) = (Z, C_-A^{-1}C_-(W))\end{aligned}$$

so that $A^* = C_-A^{-1}C_-$. Consequently $A^* = A$ if and only if $A = C_-A^{-1}C_-$ if and only if $A^{-1} = C_-AC_-$. In particular, this gives $\phi_1 = \phi_1{}^*$; that is, ϕ_1 is a Hermitian linear transformation so has real characteristic roots and is diagonalizable by a unitary transformation. If we regard ϕ_1 as a complex matrix where the adjoint matrix is the transpose of its complex conjugate, then we can write $\phi_1 = PDP^{-1}$ where D is a real diagonal matrix and $P^* = \bar{P}^t = P^{-1}$. Now set $\psi = P\tilde{D}P^{-1}$ where $\tilde{D}$ is the diagonal matrix with entries equal to the positive square root of the absolute values of the corresponding diagonal entries of D^{-1}. Note by construction that $\psi^* = \psi$.

Using this we see $\psi\phi_1 = \phi_1\psi$, $\psi^4\phi_1{}^2 = I$, and $C_-\psi^{-1}C_- = \psi = \psi^*$; that $C_-\psi^{-1}C_- = \psi$ is just a matrix computation using $C^-\phi_1^{-1}C_- = \phi_1$. Finally ψ is an automorphism as follows. Let $\lambda_1, \ldots, \lambda_m$ be the characteristic roots of ϕ which give the matrix D, and let $\tilde{g}_{\lambda_1}, \ldots, \tilde{g}_{\lambda_m}$ be the characteristic spaces. Let $X \in \tilde{g}_{\lambda_i}$ and $Y \in \tilde{g}_{\lambda_j}$. Then $[XY] = Z \in \tilde{g}_{\lambda_i\lambda_j}$ if $\lambda_i\lambda_j$ is a characteristic root, otherwise $[XY] = 0$. Now the spaces g_{λ_i} are the same characteristic spaces which give the matrix $\tilde{D}$ for ψ, and if $[XY] = Z$ as above, then

$$\begin{aligned}[\psi X\psi Y] &= |\lambda_i|^{-1/2}|\lambda_j|^{-1/2}Z \\ &= |\lambda_i\lambda_j|^{-1/2}Z = \psi Z = \psi[XY]\end{aligned}$$

which yields $\psi \in \mathrm{Aut}(\tilde{g})$.

Now let $\phi = \psi\phi_1 C_-\psi^{-1}C_- = \psi^2\phi_1$. Then $C_-\phi C_- = \phi^{-1} = \phi$ since $\phi^2 = I$. Thus, by Proposition 15.3(c), $\phi_1 C_-$ is equivalent to ϕC_- with ϕ having the required properties. Note from the construction of $\phi = \psi^2\phi_1$ we have $\phi^* = \phi$ relative to the form (Z_1, Z_2), and therefore ϕ is diagonalizable.

(b) Let K_- be a conjugation corresponding to another compact real form h_- of $\tilde{g}$. Then up to equivalence, we can write $K_- = \phi C_-$ where $\phi^2 = I$, $\phi C_- = C_-\phi$, and $\phi^* = \phi \in \mathrm{Aut}(\tilde{g})$ as above. We also have from these that $\phi = C_-K_- = K_-C_-$. Let $\langle Z_1, Z_2\rangle = -\mathrm{Kill}(Z_1, K_-(Z_2))$ for $Z_j \in \tilde{g} = h_- + ih_-$. Then as for formula (*) in part (a),

$$\mathrm{Kill}(K_-(Z_1), K_-(Z_2)) = \overline{\mathrm{Kill}(Z_1, Z_2)} = \mathrm{Kill}(C_-(Z_1), C_-(Z_2)) \qquad (**)$$

and also $\langle Z_1, Z_2\rangle$ is a positive definite Hermitian form.

Next ϕ is positive definite, for if $\phi X = \lambda X$ with $\lambda < 0$, then using (**)

$$\begin{aligned}0 > \lambda\langle X, X\rangle &= \langle \phi X, X\rangle \\ &= -\mathrm{Kill}(K_-(C_-(X)), K_-(X)) \\ &= -\mathrm{Kill}(C_-(C_-(X)), C_-(X)) = (X, X)\end{aligned}$$

which contradicts (Z_1, Z_2) is positive definite. Since ϕ is diagonalizable with real positive characteristic roots and $\phi^2 = I$ we see $\phi = I$. Thus, since we have written $K_- = \phi C_-$ up to equivalence, we see K_- is equivalent to C_-.

REMARK (3) We use the preceding results to classify the real forms of a simple complex Lie algebra $\tilde{g}$ as follows. The real forms are determined by conjugations with isomorphic forms giving equivalent conjugations. Thus we must choose a suitable representative of the equivalence class which gives isomorphic real forms. From Propositions 15.5 and 15.6 we can choose a conjugation C_- corresponding to a compact real form, and it does not make any difference (up to equivalence) what compact form is chosen. However, any other conjugation C can be written $C = \phi C_-$ with the automorphism ϕ given according to Proposition 15.6(a). Thus it suffices to discuss such automorphisms which we do for those algebras whose automorphism groups were computed in Chapter 14.

Exercises (1) Let g_- be a compact real form of a semisimple complex Lie algebra $\tilde{g}$. Let $S \in \mathrm{Aut}(\tilde{g})$ be such that $S^2 = I$, and let $k = \{X \in \tilde{g} : SX = X\}$ and $p = \{X \in \tilde{g} : SX = -X\}$. Show that the direct sum $k + ip$ is a real form of $\tilde{g}$ and every real form of $\tilde{g}$ is obtained this way up to an automorphism of $\tilde{g}$. Also note the book by Helgason [1962, pp. 152–159].

(2) Let $\tilde{g}$ be a simple complex Lie algebra, and let h be a maximal compact subalgebra of $\tilde{g}$ [note Section 12.6, remark (2)]. Show that h is a real form of $\tilde{g}$.

Proposition 15.7 Let $\tilde{g} = sl(n+1, C)$ for $n \geq 2$ be the complex Lie algebra of type A_n. Then $C_- : X \to -\bar{X}^t$ is a conjugation of a compact real form. Any given conjugation of $\tilde{g}$ is equivalent to precisely one of the following conjugations:

(a) C_-;

(b) $C_+ : X \to \bar{X}$;

(c) $C_k : X \to -T_k \bar{X}^t T_k$ for $k = 1, 2, \ldots, [(n+1)/2]$ where T_k is a diagonal matrix with the first $n + 1 - k$ entries down the diagonal equal to 1 and the last k equal to -1.

(d) $C_0 : X \to T\bar{X}T$ if $n + 1$ is even, where T is a matrix with all entries 0 except for a series of the matrices $\left[\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right]$ down its diagonal.

PROOF Since the real form of C_- is compact [example (1)], every conjugation of $\tilde{g}$ is equivalent to some φC_- where φ is an automorphism of $\tilde{g}$, $\varphi^2 = I$, and $C_- \varphi C_- = \varphi$.

From Proposition 14.22, the automorphisms of $\tilde{g}$ are given by

$$\sigma(A) : X \to AXA^{-1} \qquad \text{and} \qquad \tau(A) : X \to -AX^tA^{-1}$$

where $A \in SL(n+1, C)$. One can compute that $\sigma(A)\sigma(B) = \sigma(AB)$, $\sigma(A)\tau(B) = \tau(AB)$, $\tau(A)\sigma(B) = \tau(A((B^t)^{-1}))$, $\tau(A)\tau(B) = \sigma(A((B^t)^{-1}))$, $\sigma(A)^{-1} = \sigma(A^{-1})$, $\tau(A)^{-1} = \tau(A^t)$, and $\sigma(A) = \sigma(B)$ or $\tau(A) = \tau(B)$ if and only if B is a scalar multiple of A.

We first show that any conjugation $\sigma(A)C_-$ is equivalent to one of the conjugations in (a) or (c). We may assume $\sigma(A)^2 = \sigma(A^2) = I$ and $C_-\sigma(A)C_- = \sigma((\bar{A}^t)^{-1}) = \sigma(A)$ noting the proof of Theorem 15.6(a). Thus $A^2 = aI$ and $A\bar{A}^t = bI$ for some complex numbers a and b. Since det $A = 1$ we have $a^{n+1} = b^{n+1} = 1$. By replacing A with cA where c is a square root of $\bar{a}$, we may assume $A^2 = I$ and $A\bar{A}^t = bI$. Now $(n+1)b = \text{trace}(A\bar{A}^t) > 0$ so $b = 1$ and $\bar{A}^t = A^{-1} = A$. Thus A is a Hermitian matrix. By possibly replacing A with $-A$, we may assume that A has $+1$ and -1 as its only eigenvalues with at least as many $+1$'s as -1's. Moreover since A is Hermitian, it is diagonalizable by a unitary matrix, so we can choose a matrix B with $\bar{B}^t = B^{-1}$ and BAB^{-1} equal to I or one of the matrices T_k of (c). Finally

$$\sigma(B)\sigma(A)C_-\sigma(B)^{-1}C_- = \sigma(BA\bar{B}^t) = \sigma(BAB^{-1}),$$

and so $\sigma(A)C_-$ is equivalent to $\sigma(BAB^{-1})C_-$ which must be one of the conjugations in (a) or (c).

Next consider the conjugations of the type $\tau(A)C_-$. Assuming $I = \tau(A)^2 = \sigma(A(A^t)^{-1})$ and $C_-\tau(A)C_- = \tau((\bar{A}^t)^{-1}) = \tau(A)$, we find that $A(A^t)^{-1} = aI$ and $A\bar{A}^t = bI$. As in the first case we find that $b = 1$ so $(A^t)^{-1} = \bar{A}$ and $A\bar{A} = aI = \bar{A}A$. Now $\bar{a}I = A\bar{A} = \bar{A}A = aI$ so a is real and by taking determinants $a^{n+1} = 1$ recalling $A \in SL(n+1, C)$. There are two possibilities either $a = 1$ or if $n+1$ is even we could have $a = -1$.

So consider $\tau(A)C_-$ with $A\bar{A} = I$ and $\bar{A}^t = A^{-1}$. We claim each such conjugation is equivalent to $\tau(I)C_- : X \to \bar{X}$, the conjugation of (b). Since

$$\sigma(B)\tau(A)C_-\sigma(B)^{-1}C_- = \tau(BA\bar{B}^{-1}),$$

it suffices [by Proposition 15.3(c)] to show that there exists a $B \in SL(n+1, C)$ with

$$BA\bar{B}^{-1} = I$$

or equivalently $B^{-1}\bar{B} = A$. Since A is unitary its eigenvalues have absolute value 1 and it is diagonalizable. Thus we need only define the action of B on eigenvectors of A in order to determine B completely. If X is an eigenvector of A with corresponding eigenvalue α, then define $BX = \beta X$ where $\beta^2 = \bar{\alpha}$; note this gives $\beta\bar{\beta} = 1$. Since

$$A\bar{X} = \bar{A}^{-1}\bar{X} = \bar{\alpha}^{-1}\bar{X} = \alpha\bar{X},$$

$\bar{X}$ is also an eigenvector of A with eigenvalue α, and we can assume that we have $B\bar{X} = \beta\bar{X}$. Then

$$B^{-1}\bar{B}X = B^{-1}\bar{\beta}X = \beta^{-1}\bar{\beta}X = \bar{\beta}^2X = \alpha X = AX$$

and $B^{-1}\bar{B} = A$ as required.

The last case to be considered is $\tau(A)C_-$ with $A\bar{A} = -I$, $\bar{A}^t = A^{-1}$, and $n + 1$ is even. We shall proceed precisely as above in order to show that each of these conjugations is equivalent to $\tau(T)C_-$ where T is as in (d). We need to define B on eigenvectors of A in such a way that $BA\bar{B}^{-1} = T$. If $AX = \alpha X$, then $A\bar{X} = -\bar{A}^{-1}\bar{X} = -\bar{\alpha}^{-1}\bar{X} = -\alpha\bar{X}$, so eigenvectors of A occur in pairs $X, \bar{X}$ with corresponding eigenvalues the negatives of each other. If $AX = \alpha X$ define $B_1X = \beta X$ where $\beta^2 = \bar{\alpha}$. Then clearly we can assume $B_1\bar{X} = \beta\bar{X}$ so that $\bar{B}_1^{-1}X = \overline{B_1^{-1}\bar{X}} = \overline{\beta^{-1}\bar{X}} = \beta X$, $B_1A\bar{B}_1^{-1}X = \beta^2\alpha X = X$, and $B_1A\bar{B}_1^{-1}\bar{X} = -\beta^2\alpha\bar{X} = -\bar{X}$. Using the fact that $A' = B_1A\bar{B}_1^{-1}$ has half of its eigenvalues equal to $+1$ and the other half equal to -1, A' has the same Jordan canonical form as T and consequently $B_2A'B_2^{-1} = T$ for some B_2. Next from the form of T notice that iT is a real matrix. Also for eigenvectors X, $\bar{X}$ as above we see $A'X = X$ and $A'\bar{X} = -\bar{X}$ implies $\bar{A}'\bar{X} = \bar{X}$ and $\bar{A}'X = -X$ which gives $\bar{A}' = -A'$. Therefore $\overline{iA'} = \bar{i}\bar{A}' = (-i)(-A') = iA'$ so that iA' is also a real matrix. This gives for the real matrix $iT = B_2(iA')B_2^{-1}$. Thus the real matrices iT and iA' can be assumed to be similar by a real matrix B_2 since they are similar by a complex matrix. We can conclude $T = B_2A'B_2^{-1}$ with B_2 a real matrix. However, then $\bar{B}_2^{-1} = B_2^{-1}$ and for $B = B_2B_1$ we see

$$BA\bar{B}^{-1} = (B_2B_1)A(\overline{B_2B_1})^{-1} = B_2A'B_2^{-1} = T$$

as required.

The fact that none of the conjugations listed are equivalent follows from the fact that their real forms are not isomorphic, and in fact all have distinct signatures. The signatures are listed in the following corollary, and showing that the list is correct is left as an exercise.

Corollary 15.8 Any real form of $\tilde{g} = sl(n + 1, C)$ of type A_n, $n \geq 2$ must be isomorphic to precisely one of the algebras shown in Table 15.2.

Exercises (3) Use remark (2) of this section to verify that the correct signatures are given in the above tabulation. Also show that none of the signatures can coincide for a fixed n.

(4) Show that the 2×2 matrices X of complex numbers such that $T_0\bar{X}T_0 = X$ where $T_0 = \left[\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right]$ form an algebra over the real number which

TABLE 15.2

	Name	Signature
(a)	$su(n+1) = \{X \in \tilde{g} : \bar{X}^t = -X\}$	$-n^2 - 2n$
(b)	$sl(n+1, R) = \{X \in \tilde{g} : \bar{X} = X\}$	n
(c)	$\mathscr{A}(n, k, R) = \{X \in \tilde{g} : -T_k \bar{X}^t T_k = X\}$ for $k = 1, 2, \ldots, [(n+1)/2]$ and T_k as in Proposition 15.7.	$1 - (n + 1 - 2k)^2$
(d)	If $n + 1$ is even, $sl(n + 1/2, \mathscr{Q})$ is the set of $(n+1)/2 \times (n+1)/2$ matrices with entries from $\mathscr{Q}$, the real quaternion numbers, such that the trace of the matrices is a quaternion number of trace 0.	$-n - 2$

is isomorphic to the quaternion numbers. Also show that $sl((n+1)/2, \mathscr{Q})$ is isomorphic to $g_0 = \{X \in \tilde{g} : T\bar{X}T = X\}$ where $n + 1$ is even, $\tilde{g} = sl(n+1, C)$, and T is as in Proportion 15.7(d).

(5) Show that any real form of $\tilde{g} = sl(2, C)$ of type A_1 must be isomorphic to $sl(2, R)$ or $su(2)$.

Proposition 15.9 Let $\tilde{g}$ be a simple complex Lie algebra of type G_2. Then there are only two isomorphism classes of real forms of $\tilde{g}$, the split real forms and the compact real forms.

PROOF We may assume $\tilde{g} = \mathscr{D}(\mathscr{C})$, the derivation algebra of $\mathscr{C}$ the complex Cayley algebra as described in example (2) of Section 9.6. Let $\lambda : \mathscr{C} \to \mathscr{C}$ denote the map which sends an element $\left[\begin{smallmatrix} a & u \\ v & b \end{smallmatrix}\right]$ into $\left[\begin{smallmatrix} \bar{a} & \bar{u} \\ \bar{v} & \bar{b} \end{smallmatrix}\right]$ where these elements were defined in the example mentioned above. Clearly $\lambda^2 = I$, $\lambda(aX) = \bar{a}\lambda(X)$ where $a \in C$ and $X \in \mathscr{C}$, and $\mathscr{C}_+ = \{X \in \mathscr{C} : \lambda(X) = X\}$ is a split real Cayley algebra. Define $C_+ : \tilde{g} \to \tilde{g}$ by $C_+ : D \to \lambda D \lambda$. Then it is easy to verify that C_+ is a conjugation of $\tilde{g}$ and its real form is $\mathscr{D}(\mathscr{C}_+)$, a split algebra of type G_2. By following the similar arguments for Lie algebras, it can be shown that every real Cayley algebra is the fixed point set of some $T\lambda$ with $T \in \mathrm{Aut}(\mathscr{C})$, and the real algebras $\mathscr{C}_1$ and $\mathscr{C}_2$ corresponding to $T_1\lambda$ and $T_2\lambda$ are isomorphic if and only if there is some $T \in \mathrm{Aut}\,\mathscr{C}$ with $T_2 = TT_1\lambda T^{-1}\lambda$. However, it follows from Proposition 14.23 that $\mathrm{Aut}(\tilde{g})$ is equal to the set of automorphisms $\varphi(T) : D \to TDT^{-1}$ with $T \in \mathrm{Aut}(\mathscr{C})$. Finally one can check that the real form of $\varphi(T_1)\mathscr{C}_+$ is $\mathscr{D}(\mathscr{C}_1)$ where $\mathscr{C}_1$ is the fixed point set of $T\lambda$ and that $\mathscr{D}(\mathscr{C}_1)$ and $\mathscr{D}(\mathscr{C}_2)$ are isomorphic if and only if $\mathscr{C}_1$ and $\mathscr{C}_2$ are isomorphic. Since we have mentioned that there are only two isomorphism classes of real Cayley algebras, the same must be true for their derivation algebras.

Proposition 15.10 Let $\tilde{g} = \mathscr{D}(n, C)$ be the simple complex Lie algebra of type D_n and assume $n \geq 5$; that is, $\tilde{g} = \{X \in gl(2n, C) : LX^tL = -X\}$ where

$$L = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}.$$

The following is a listing of one representative from each isomorphism class of real forms of $\tilde{g}$.

(a) $\mathscr{D}(n, k, R) = \{X \in \tilde{g} : -T_k\bar{X}^tT_k = X\}$ where $k = 0, 1, 2, \ldots, n$ and

$$T_k = \left[\begin{array}{c|c|c|c} I & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & I \\ \hline 0 & 0 & I & 0 \\ \hline 0 & I & 0 & 0 \end{array}\right]\begin{array}{l} \}\,n-k \\ \}\,k \\ \}\,n-k \\ \}\,k \end{array}.$$

The signature of $\mathscr{D}(n, k, R)$ is $2k(2n - k) - n(2n - 1)$, and for $k = 0$ the form is compact.

(b) $\mathscr{D}(n, -1, R) = \{X \in \tilde{g} : -T_{-1}\bar{X}^tT_{-1}^{-1} = X\}$ where

$$T_{-1} = i\left[\begin{array}{c|c} I & 0 \\ \hline 0 & -I \end{array}\right]\begin{array}{l} \}\,n \\ \}\,n. \end{array}$$

The signature of $\mathscr{D}(n, -1, R)$ is $-n$.

Proof We must first describe $\mathrm{Aut}(\tilde{g})$. It was seen in Chapter 14 that the inner automorphisms of $\tilde{g}$ consists of all $\varphi(T) : X \to TXT^{-1}$ where $LTL = T^{-1}$ and $\det(T) = 1$. It is clear that $\varphi(T)$ is an automorphism even if $\det(T) = -1$, and it is claimed that such $\varphi(T)$'s are not inner automorphisms. Thus $\mathrm{Aug}(\tilde{g}) = \{\varphi(T) : LTL = T^{-1}$ and $\det T = \pm 1\}$. To show that we cannot have $\varphi(T_2) = \varphi(T_1)$ with $\det(T_2) = -1$ and $\det(T_1) = +1$, merely notice that this would imply $\varphi(T_1^{-1}T_2) = I$ so $T_1^{-1}T_2 = \pm I$ and $\det(\pm I) = 1 \neq -1 = \det(T_1^{-1}T_2)$.

It is clear that $\mathscr{D}(m, 0, R) = \{X \in \tilde{g} : -\bar{X}^t = X\}$ is a compact real form so denote its conjugation by $C_- : X \to -\bar{X}^t$. Every conjugation of $\tilde{g}$ is equivalent to one of the type $\varphi(T)C_-$ where $\varphi(T)^2 = \varphi(T^2) = I$ and $C_-\varphi(T)C_- = \varphi((\bar{T}^t)^{-1}) = \varphi(T)$. The first condition implies $T^2 = \pm I$ and the second implies $T\bar{T}^t = I$.

We first consider the case $T^2 = I$. Then T has eigenvalues ± 1 and we can assume that the $+1$ eigenspace has at least the dimension of the -1 eigenspace. We can consider the nondegenerate symmetric bilinear form on the $2n$-dimensional complex vector space V of column vectors defined by $(x, y) = x^tLy$ so $\tilde{g} = \{X \in gl(V) : (Xx, y) + (x, Xy) = 0\}$ and $LT^tL = T^{-1}$ if and only if $(Tx, Ty) = (x, y)$. Now with T as above, suppose $Tx = x$ and $Ty = -y$. Then $(x, y) = (Tx, y) = (x, T^{-1}y) = -(x, y)$ so the $+1$ eigenspace and -1 eigenspace are orthogonal with respect to (x, y). Next notice that for each T_k of (a) we have $T_k^2 = I$, $T_k \bar{T}_k^t = I$ and the $+1$ eigenspace is orthogonal to the -1 eigenspace for T_k. It is also easy to verify that the -1 eigenspace of T_k has dimension k, so it follows that there exists an S with $LS^tL = S^{-1}$ and for some k, $STS^{-1} = T_k$. Since $\bar{T}^t = T^{-1} = T$ and T is Hermitian, we can assume that S is unitary; that is, $\bar{S}^t = S^{-1}$. However, then noting

$$\begin{aligned} C_-\phi(S)^{-1}C_-(X) &= C_-\phi(S)^{-1}(-\bar{X}^t) \\ &= C_-(-S^{-1}\bar{X}^tS) = \bar{S}^tX(\bar{S}^t)^{-1} = \phi(\bar{S}^t)(X), \end{aligned}$$

we obtain $\phi(S)\phi(T)C_-\phi(S)^{-1}C_- = \phi(ST\bar{S}^t) = \phi(STS^{-1}) = \phi(T_k)$. Therefore by Proposition 15.3(c), $\phi(T)C_-$ is equivalent to $\phi(T_k)C_-$.

For the second case $T^2 = -I$, we repeat the now familiar agruments once more. Thus T has eigenvalues $\pm i$ and if $Tx = ix$, $Ty = iy$, then $i(x, y) = (Tx, y) = (x, T^{-1}y) = -i(x, y)$, so $(x, y) = 0$ for all x, y in the $+i$ eigenspace. This implies the $+i$ and $-i$ eigenspaces have the same dimension. For T_{-1} as in statement (b) we see $T_{-1}^2 = -I$ and T_{-1} has eigenspaces like T, so once again we can show that $\varphi(T)C_-$ and $\varphi(T_{-1})C_-$ are equivalent.

It remains to show that none of the algebras on our list are isomorphic. First we would have to verify that the correct signatures are listed and then check that this helps imply the desired conclusion. The calculation of the signatures will be omitted since they are straightforward but long and cumbersome. The remainder of the calculations are given in the following exercise.

Exercise (6) Assume the signatures as given in the previous proposition. Then show that for a given n the only algebras which have the same signatures are $\mathscr{D}(n, n - n^{1/2}, R)$ and $\mathscr{D}(n, -1, R)$ if n happens to be a perfect square. Show that these two algebras are not isomorphic.

Proposition 15.11 Let $\tilde{g} = \mathscr{B}(n, C)$ with $n \geq 2$ as described in exercise (1), Section 13.2. Then the following list gives one representative from each isomorphism class of real forms: $\mathscr{B}(n, k, R) = \{X \in \tilde{g} : T_k'\bar{X}^tT_k' = X\}$ where

$$T_k' = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & T_k \end{array}\right]$$

for $k = 0, 1, 2, \ldots, n$ and T_k is as in Proposition 15.10. The signature of $\mathscr{B}(n, k, R)$ is $(2k - n)(2n + 1) - 2k^2$, and for $k = 0$ the form is compact.

Proposition 15.12 Let $\tilde{g} = \mathscr{C}(n, C)$ with $n \geq 3$. Then the following list gives one representative for each isomorphism class of real forms of $\tilde{g}$.

(a) $\mathscr{C}(n, k, R) = \{X \in \tilde{g} : -S_k X^t S_k = X\}$ for $k = 0, 1, 2, \ldots, [n/2]$ and

$$S_k = \begin{array}{c} \left[\begin{array}{c|c|c|c} I & 0 & 0 & 0 \\ \hline 0 & -I & 0 & 0 \\ \hline 0 & 0 & I & 0 \\ \hline 0 & 0 & 0 & -I \end{array}\right] \end{array} \begin{array}{l} \}\, n-k \\ \}\, k \\ \}\, n-k \\ \}\, k \end{array}$$

The signature of $\mathscr{C}(m, k, R)$ is $8k(n-k) - n(2n + 1)$, and for $k = 0$ the form is compact.

(b) $\mathscr{C}(n, -1, R) = \{X \in \tilde{g} : -T_{-1} X^t T_{-1}^{-1} = X\}$ where T_{-1} is as in Proposition 15.10. The signature of $\mathscr{C}(n, -1, R)$ is n.

Exercises (7) Show that a proof of Proposition 15.11 is actually contained in the proof of Proposition 15.10.

(8) Prove Proposition 15.12.

REMARK (4) We do not have enough information about the automorphism groups of the complex Lie algebras D_4, F_4, E_6, E_7, and E_8 to classify their real forms in the same manner as the others. We list the number of real forms of each type and their signatures in Table 15.3 without proof. A very explicit proof of all of these facts can be found in the work of Freudenthal and De Vries [1969].

TABLE 15.3

Type	Number of classes of real forms	Signatures
D_4	5	4, 2, −4, −14, −28
F_4	3	4, −20, −52
E_6	5	6, 2, −14, −26, −78
E_7	4	7, −5, −25, −133
E_8	3	8, −24, −248

2. Representations of Real and Complex Simple Lie Algebras

Since the simple complex Lie algebras and their real forms are all completely reducible, the finite-dimensional representations of these algebras are all direct sums of irreducible ones. In this section we will describe the irreducible representations of the simple complex Lie algebras and their relationship to the representations of the real forms.

REMARK (1) For the remainder of this section $\tilde{g}$ will denote a simple complex Lie algebra, $\tilde{h}$ a Cartan subalgebra of $\tilde{g}$, $\mathscr{R}$ the set of nonzero roots of $\tilde{g}$, and $\mathscr{B}$ a basis for $\mathscr{R}$. We wish to determine the equivalence classes of finite-dimensional irreducible representations of $\tilde{g}$ where two representations $\zeta_i : \tilde{g} \to gl(V_i)$, $i = 1, 2$, are called **equivalent representations** if there exists a nonsingular linear transformation T: $V_1 \to V_2$ with $\zeta_2(X) = T\zeta_1(X)T^{-1}$ for all $X \in \tilde{g}$. As in the case of the adjoint representation, if $\zeta : \tilde{g} \to gl(V)$ is a representation of $\tilde{g}$, then V is a direct sum of weight spaces of $\zeta(\tilde{h})$. More explicitly let $V(\lambda) = \{v \in V : (\zeta(H) - \lambda(H)I)^k v = 0$ for all $H \in \tilde{h}$ and some positive integer $k\}$. Here λ is any linear transformation $\lambda : \tilde{h} \to C$ and λ is called a **weight** of ζ if $V(\lambda)$ is nonzero. Let $\mathscr{W}$ denote the set of all weights of ζ (including 0 perhaps). Then

$$V = \sum_{\lambda \in \mathscr{W}} V(\lambda)$$

as a vector space direct sum.

Proposition 15.13 (a) If $\alpha \in \mathscr{R}$, $X \in \tilde{g}(\alpha)$, $\lambda \in \mathscr{W}$, and $v \in V(\lambda)$, then $\zeta(X)(v) \in V(\alpha + \lambda)$ if $\alpha + \lambda \in \mathscr{W}$ and otherwise $\zeta(X)(v) = 0$.

(b) The weight spaces are in fact eigenspaces; that is, $\zeta(H)(v) = \lambda(H)v$ for all $H \in \tilde{h}$ and $v \in V(\lambda)$.

(c) If $\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, then there exists a unique weight $\lambda_0 \in \mathscr{W}$ such that $\lambda_0 + \alpha_k \notin \mathscr{W}$ for $k = 1, 2, \ldots, n$, and λ_0 is called the **maximal weight** of ζ.

(d) If $\lambda \in \mathscr{W}$, then $\lambda = \lambda_0 - \sum_{k=1}^n t_k \alpha_k$ for some positive integers t_k.

(e) The dimension of $V(\lambda_0)$ is 1.

PROOF (a) and (b) Clearly there exists some $\lambda \in \mathscr{W}$ and $v \in V(\lambda)$ such that $\zeta(H)(v) = \lambda(H)v$ for all $H \in \tilde{h}$. Then for any $X \in \tilde{g}(\alpha)$

$$\begin{aligned}\zeta(H)\zeta(X)(v) &= \zeta(X)\zeta(H)(v) + [\zeta(H), \zeta(X)](v)\\ &= (\lambda(H) + \alpha(H))\zeta(X)(v)\end{aligned}$$

so either $\zeta(X)(v) = 0$ or $\zeta(X)(v)$ is an eigenvector of $\zeta(H)$ and $\zeta(X)(v) \in V(\alpha + \lambda)$. Now it is clear that v generates a $\tilde{g}$-submodule of V which consists entirely of eigenvectors of $\zeta(\tilde{h})$. Since V is irreducible this $\tilde{g}$-submodule must be all of V.

(c)–(e) Let $\lambda_0 \in \mathscr{W}$ be a weight such that $\lambda_0 + \alpha_k \notin \mathscr{W}$ for $k = 1, 2, \ldots, n$. Such a weight must exist because V is finite dimensional and so has only a finite number of weights, and given any weight we can continue to add α_k's to this weight until it satisfies this property.

Let $0 \neq X_k \in \tilde{g}(\alpha_k)$ and $0 \neq Y_k \in \tilde{g}(-\alpha_k)$ for $k = 1, 2, \ldots, n$. Choose a fixed $v_0 \in V(\lambda_0)$. Then since V is irreducible, it is clear that V is spanned by vectors of the type $T(v_0)$ where T is a product of certain $\zeta(X_k)$'s and $\zeta(Y_k)$'s. Also notice that $\lambda_0 + \alpha_k \notin \mathscr{W}$ implies that $\zeta(X_k)(v_0) = 0$ for $k = 1, 2, \ldots, n$. We claim that V is spanned by a vector of the type $T(v_0)$ where T is a product of certain $\zeta(Y_k)$'s. To see this set $H_k = [X_k, Y_k]$ so that $H_k \in \tilde{h}$ and notice that if $j \neq k$, then using Proposition 14.17

$$\zeta(X_j)\zeta(Y_k)(v_0) = \zeta([X_j Y_k])(v_0) + \zeta(Y_k)\zeta(X_j)(v_0) = 0$$

and

$$\zeta(X_k)\zeta(Y_k)(v_0) = \zeta(H_k)(v_0) + \zeta(Y_k)\zeta(X_k)(v_0) = \lambda_0(H_k)v_0.$$

By making repeated use of the above procedure we see that we can eliminate any $\zeta(X_k)$'s from longer products.

Now suppose that $T(v_0) \neq 0$, where T is a certain product of $\zeta(Y_k)$'s where $\zeta(Y_k)$ occurs t_k times. Then $T(v_0) \in V(\lambda)$ where

$$\lambda = \lambda_0 - \sum_{k=1}^{n} t_k \alpha_k.$$

This proves (d).

To see that the dimension of $V(\lambda_0)$ is 1, notice that the formula above guarantees that none of $T(v_0)$'s, with T a nontrivial product of $\zeta(Y_k)$'s, are in $V(\lambda_0)$. Thus $V(\lambda_0)$ consists of all multiples of v_0 and the proof of (e) is completed.

Finally we must show that λ_0 is the unique maximal weight. Our proof has shown that all weights of ζ can be found from a maximal weight by the formula in (d) and it is clear that only one weight can satisfy this property.

REMARKS (2) Proposition 15.13 implies in particular that each root system of a simple complex Lie algebra has a unique root of maximal height; that is, there is a unique root $\sum_{k=1}^{n} t_k \alpha_k \in \mathscr{R}^+$ with $\sum_{k=1}^{n} t_k$ maximal.

(3) Although $\dim V(\lambda_0) = 1$ we may have $\dim V(\lambda) > 1$ for other $\lambda \in \mathscr{W}$. In fact this is always true for the weight 0 of the adjoint representation of a $\tilde{g}$ with rank $\tilde{g} > 1$.

Proposition 15.14 For any $\alpha \in \mathscr{R}$ let $H_\alpha' \in \tilde{h}$ be as usual so that we can choose $X \in \tilde{g}(\alpha)$, $Y \in \tilde{g}(-\alpha)$ with $[H_\alpha' X] = 2X$, $[H_\alpha' Y] = -2Y$, and $[XY] = H_\alpha'$. Then $\lambda(H_\alpha')$ is an integer for each $\lambda \in \mathscr{W}$.

Proof Let g' be the subalgebra of $\tilde{g}$ generated by H_α', X and Y so that $g' \cong sl(2, C)$. Choose any $v \in V(\lambda)$. Then the action of g' on v generates an irreducible g' subspace V' of V. Now $\zeta(H_\alpha')(v) = \lambda(H_\alpha')v$ and $\lambda(H_\alpha')$ must be an integer by Theorem 13.11.

Example (1) Let $\tilde{g} = sl(2, C)$. Then for $\mathscr{R} = \{-\alpha, \alpha\}$ we have $H_\alpha' = \left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$. The weights of the adjoint representation are $\mathscr{W}_1 = \{-\alpha, 0, \alpha\}$ and $\alpha(H_\alpha') = 2$. Let V be the vector space of two-dimensional complex column vectors with $\tilde{g}$ acting on V as usual by matrix multiplication. Then the weights of this representation are $\mathscr{W}_2 = \{-\alpha/2, \alpha/2\}$ and $V(\alpha/2)$ is spanned by the vector $\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]$. Finally $\frac{1}{2}\alpha(H_\alpha') = 1$ an integer as guaranteed by the proposition.

Theorem 15.15 Let $\tilde{g}$ be a simple complex Lie algebra and $\lambda_0 : \tilde{h} \to C$ a linear transformation such that each $\lambda_0(H'_{\alpha_k})$ is a positive integer or 0 for $\alpha_k \in \mathscr{B}$. Then there exists an irreducible representation $\zeta : \tilde{g} \to gl(V)$ with λ_0 as maximal weight. The representation is finite-dimensional, determined up to equivalence by λ_0 and every finite-dimensional representation of $\tilde{g}$ is equivalent to some such ζ.

Proof It is clear from the proof of Proposition 15.14 that $\lambda_0(H'_{\alpha_k})$ is a nonnegative integer if λ_0 is the maximal weight of a finite-dimensional representation of $\tilde{g}$. Therefore, all we need show is that ζ is completely determined once we are given some $0 \neq v_0 \in V(\lambda_0)$, $\tilde{g}$ and the nonnegative integers $\lambda_0(H'_{\alpha_k}) = a_k$.

Now for $k = 1, 2, \ldots, n$, let $H_k = H'_{\alpha_k}$ and choose $X_k \in \tilde{g}(\alpha_k)$, $Y_k \in g(-\alpha_k)$ so that H_k, X_k, Y_k forms the usual basis for a subalgebra g_k of $\tilde{g}$ with $g_k \cong sl(2, C)$. By Theorem 13.11 the number a_k completely determines the action of g_k on an irreducible subspace of V with v_0 in the weight space of the maximal weight. More precisely $\zeta(Y_k)^j(v_0)$, $j = 0, 1, 2, \ldots, p$, for some p is a basis for one of the modules described in Theorem 13.11.

It suffices to show that we can determine the action of all of $\tilde{g}$ on basis elements of the type

$$v = \zeta(Y_{k_1})^{t_1} \cdots \zeta(Y_{k_q})^{t_q}(v_0)$$

and we proceed by induction on $\sum_{j=1}^{q} t_j$. Since $\zeta(X_k)\zeta(Y_k) = \zeta(Y_k)\zeta(X_k) + \zeta(H_k)$, $\zeta(X_j)\zeta(Y_k) = \zeta(Y_k)\zeta(X_j)$ if $j \neq k$ and $\zeta(X_k)(v_0) = 0$, we can determine the action of $\zeta(X)$ on v for all $X \in \tilde{g}(\alpha)$ with $\alpha \in \mathscr{R}^+$. For any $H \in \tilde{h}$

$$\zeta(H)(v) = (\lambda_0(H) - \sum_{j=1}^{q} t_j \alpha_{k_j}(H))v.$$

Next to determine whether or not $\zeta(Y_k)(v)$ is 0 we first set

$$v_1 = \zeta(X_k)^r(v) \neq 0 \qquad \text{where} \quad \zeta(X_k)^{r+1}(v) = 0.$$

Then the integer $\zeta(H_k)(v_1)$ determines an irreducible representation of g_k with v_1 as maximal weight vector. From this we determine whether or not $\zeta(Y_k)(v)$ is 0.

To show that the module generated is finite dimensional we need only show that for each $k = 1, \ldots, n$ there are only finitely many irreducible g_k submodules of our $\tilde{g}$-module. From the fact that each weight space of our module is finite dimensional and there are only a finite number of weights λ of our representation with $\lambda(H_k) > 0$, we conclude there are only finitely many maximal weight spaces for irreducible representations of g_k.

Finally we must determine a basis for each weight space.

Exercise (1) Finish the proof of Theorem 15.15 by showing that a basis for each weight space can be determined from $\tilde{g}$ and the integers $a_1, \ldots, a_n$.

Definition 15.16 The representations ζ_k, $k = 1, \ldots, n$ of $\tilde{g}$ with maximal weights λ_k such that $\lambda_k(H_k) = 1$ and $\lambda_k(H_j) = 0$ for $j \neq k$ are called the **basic representations** of $\tilde{g}$.

Proposition 15.17 If μ_1 and μ_2 are finite-dimensional representations of $\tilde{g}$ with $\mu_j : \tilde{g} \to gl(V_j)$, $j = 1, 2$ and μ_j has maximal weight λ_j, then define $\mu_1 \otimes \mu_2$ on $V_1 \otimes V_2$ by

$$(\mu_1 \otimes \mu_2)(X) : u \otimes v \to \mu_1(X)(u) \otimes v + u \otimes \mu_2(X)(v).$$

Let $0 \neq v_j \in V_j(\lambda_j)$ for $j = 1, 2$, then $\mu_1 \otimes \mu_2 : \tilde{g} \to gl(V_1 \otimes V_2)$ is a representation of $\tilde{g}$ such that the submodule of $V_1 \otimes V_2$ generated by $v_1 \otimes v_2$ is the module of an irreducible representation with maximal weight $\lambda_1 + \lambda_2$.

PROOF Some trivial computations show that $\mu_1 \otimes \mu_2$ is a representation; note Sections 9.4 and 12.5. Now

$$(\mu_1 \otimes \mu_2)(H)(v_1 \otimes v_2) = (\lambda_1 + \lambda_2)(H)(v_1 \otimes v_2)$$

and clearly $v_1 \otimes v_2$ belongs to a maximal weight space of $V_1 \otimes V_2$ so the proposition follows.

Corollary 15.18 Any finite-dimensional irreducible representation of $\tilde{g}$ can be obtained from some tensor product of basic irreducible representations as in the previous proposition.

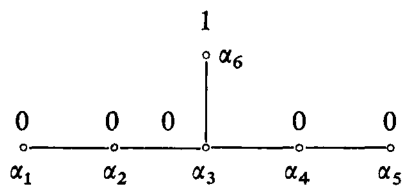
PROOF Let ζ be an irreducible representation of $\tilde{g}$ with maximal weight λ and set $a_k = \lambda(H_k)$ for $k = 1, 2, \ldots, n$. Each a_k is a nonnegative integer so we can set V equal to a tensor product of V_k's with V_k occurring a_k times where $\mu_k : \tilde{g} \to gl(V_k)$ are the basic irreducible representations defined in Definition 15.16. This tensor product yields a representation as defined in Proposition 15.17, and the tensor product of the maximal weight vectors of the V_k's generates a $\tilde{g}$-submodule of V with has the same maximal weight as ζ.

Exercises (2) Show that the adjoint representation of the complex Lie algebra of type G_2 is a basic representation. Also show that the representation of this algebra as derivations of a Cayley algebra is basic and not equivalent to the adjoint representation.

(3) Is the adjoint representation of $\tilde{g} = sl(n+1, C)$ a basic representation?

REMARK (4) Often a representation is described by adding to the Dynkin diagram of $\tilde{g}$ the number $\lambda_0(H_k)$ to the vertex of the diagram corresponding to the root α_k. Since the Dynkin diagram completely determines the structure of $\tilde{g}$ and an irreducible representation ζ is completely determined by the numbers $\lambda_0(H_k)$ for its maximal weight λ_0, theoretically one could construct the Lie algebra and its representation using only the diagram. In practice, of course, the computations are prohibitive, but it is not too difficult to determine the maximal weight λ_0 from the diagram.

Example (2) Consider the accompanying diagram with 0's and 1's attached to the roots which is the diagram for a basic irreducible represen-



tation of the complex Lie algebra of type E_6. We can set $\lambda_0 = \sum_{k=1}^{6} t_k \alpha_k$ and we would like to determine the coefficients t_k. We know

$$\lambda_0(H_6) = 2\langle \lambda_0, \alpha_6 \rangle / \langle \alpha_6, \alpha_6 \rangle = 1 \qquad \text{and} \qquad \lambda_0(H_k) = 2\langle \lambda_0, \alpha_k \rangle / \langle \alpha_k, \alpha_k \rangle = 0$$

for $k = 1, \ldots, 5$. Also

$$2\langle \alpha_k, \alpha_k \rangle / \langle \alpha_k, \alpha_k \rangle = 2, \qquad 2\langle \alpha_1, \alpha_2 \rangle / \langle \alpha_1, \alpha_1 \rangle = -1,$$
$$2\langle \alpha_1, \alpha_3 \rangle / \langle \alpha_1, \alpha_1 \rangle = 0$$

and so forth. Thus

$$\begin{aligned}
\lambda_0(H_1) &= 2t_1 - t_2 = 0,\\
\lambda_0(H_2) &= -t_1 + 2t_2 - t_3 = 0,\\
\lambda_0(H_3) &= -t_2 + 2t_3 - t_4 - t_6 = 0,\\
\lambda_0(H_4) &= -t_3 + 2t_4 - t_5 = 0,\\
\lambda_0(H_5) &= -t_4 + 2t_5 = 0,\\
\lambda_0(H_6) &= -t_3 + 2t_6 = 1.
\end{aligned}$$

Solving the system of equations one finds that $t_1 = t_5 = 1$, $t_2 = t_4 = t_6 = 2$, and $t_3 = 3$ so $\lambda_0 = \alpha_1 + 2\alpha_2 + 3\alpha_3 + 2\alpha_4 + \alpha_5 + 2\alpha_6$. This is, in fact, the maximal weight of the adjoint representation of the complex Lie algebra of type E_6. There are convenient tables listing the maximal weights of all of the basic representations for all of the simple complex Lie algebras as well as a listing of all of the weights of the adjoint representations in the appendix of the book by Freudenthal and de Vries [1969].

REMARK (5) Given an irreducible representation $\zeta : \tilde{g} \to gl(V)$ with maximal weight λ_0 and with $\mathscr{R}^+$ the positive roots of $\tilde{g}$, then the dimension of V can be obtained from **Weyl's formula**

$$\dim V = \prod_{\alpha \in \mathscr{R}} \langle \lambda_0 + \beta, \alpha \rangle / \langle \beta, \alpha \rangle$$

where $\beta = \frac{1}{2} \sum_{\alpha \in \mathscr{R}^+} \alpha$. The proof of Weyl's formula uses properties of representations we have not discussed so it will not be given here. A proof can be found in the book by Jacobson [1962].

Exercise (4) Let $\tilde{g}$ be a complex Lie algebra of type A_3 and $\zeta : \tilde{g} \to gl(V)$ a basic irreducible representation corresponding to the accompanying diagram. Assuming Weyl's formula and assuming that we know that

$$\begin{array}{ccccc} 0 & & 1 & & 0 \\ \circ & \text{———} & \circ & \text{———} & \circ \\ \alpha_1 & & \alpha_2 & & \alpha_3 \end{array}.$$

$\mathscr{R}^+ = \{\alpha_1, \alpha_2, \alpha_3, \alpha_1 + \alpha_2, \alpha_2 + \alpha_3, \alpha_1 + \alpha_2 + \alpha_3\}$, find the maximal weight λ_0 of ζ as a linear combination of the α_k's and find the dimension of V.

Examples (3) The complex Lie algebras of type A_n, B_n, C_n, and D_n were all defined as certain sets of matrices so this definition includes the description of a representation on a set of column vectors. Thus $sl(n + 1, C)$, $\mathscr{B}(n, C)$, $\mathscr{C}(n, C)$, and $\mathscr{D}(n, C)$ are descriptions of representations on spaces of dimension $n + 1$, $2n + 1$, $2n$, and $2n$, respectively. In each case these are representations on spaces of minimal dimension.

(4) Consider two irreducible representations $\zeta_k : \tilde{g} \to gl(V_k)$, $k = 1, 2$, such that the diagram for ζ_2 is obtained by permuting the numbers assigned to the vertices of the Dynkin diagram by reflecting them through a symmetry of Dynkin diagram. From the symmetries involved in constructing $\tilde{g}$ and the representations from the diagrams it is clear that $\dim V_1 = \dim V_2$ and there will be a natural correspondence between weight spaces. Thus we can assume that $V_1 = V_2 = V$ and the representations ζ_1 and ζ_2 will resemble each other even if they are not equivalent. For example, if $\tilde{g} = sl(n+1, C)$ for $n \geq 2$ (so $\tilde{g}$ is of type A_n) and we consider the representations corresponding to the accompanying diagrams. It can be shown that in this case $\dim(V) = n + 1$

$$\underset{\zeta_1}{\begin{array}{ccccc} 1 & 0 & & 0 & 0 \\ \circ\!-\!\!- & \circ\!-\!\!- & \cdots & \!-\!\!\circ\!-\!\!- & \circ \\ \alpha_1 & \alpha_2 & & \alpha_{n-1} & \alpha_n \end{array}}, \qquad \underset{\zeta_2}{\begin{array}{ccccc} 0 & 0 & & 0 & 1 \\ \circ\!-\!\!- & \circ\!-\!\!- & \cdots & \!-\!\!\circ\!-\!\!- & \circ \\ \alpha_1 & \alpha_2 & & \alpha_{n-1} & \alpha_n \end{array}}.$$

and we can let V be the vector space of complex $n + 1$-dimensional column vectors. For $A \in gl(n+1, C)$ and $x \in V$ we can set $\zeta_1(A)(x) = Ax$ the usual matrix multiplication and $\zeta_2(A)(x) = -A^t x$. The statement that ζ_1 and ζ_2 are not equivalent is the same as saying that $A \to -A^t$ is not an inner automorphism of $\tilde{g}$.

REMARK (6) We will now consider representations of real forms of a simple complex Lie algebra. Given a real form g of $\tilde{g}$ and $\tilde{\zeta} : \tilde{g} \to gl(\tilde{V})$ a representation of $\tilde{g}$, then $\tilde{\zeta}$ restricted to g is clearly a representation of g but is considered to be unsatisfactory in certain ways because this gives an action of the real algebra g on a complex vector space $\tilde{V}$. This can be remedied by restricting the scalar multiplication on $\tilde{V}$ to the real numbers, but a more useful solution is described in the next definition. Notice that we have altered our notation slightly, for the remainder of this section $\tilde{g}$, $\tilde{V}$, $\tilde{\zeta}$ will denote the complex Lie algebras, vector spaces and representations and g, V, ζ will denote real ones.

Definition 15.19 (a) Given a finite-dimensional complex vector space $\tilde{V}$ and $C : \tilde{V} \to \tilde{V}$ a nonsingular linear transformation of $\tilde{V}$ when we restrict scalar multiplication in $\tilde{V}$ to the real numbers. If $C^2 = I$ and $C(ax) = \bar{a}C(x)$ for all $a \in C$ and $x \in \tilde{V}$, then C is called a **conjugation** of $\tilde{V}$ and

$$V = \{x \in \tilde{V} : C(x) = x\}$$

is said to be the **real form** of $\tilde{V}$ corresponding to C.

(b) Given a real form g of $\tilde{g}$, a simple complex Lie algebra, and $\tilde{\zeta} : \tilde{g} \to gl(\tilde{V})$ an irreducible representation of $\tilde{g}$, then $\tilde{\zeta}$ is said to be **real for** g if there exists a conjugation C of $\tilde{V}$ with corresponding real form V such that

$\tilde{\zeta}(X)(x) \in V$ for all $X \in g$ and $x \in V$. In this case the restriction ζ of $\tilde{\zeta}$ to g acting on V is said to be a **real form of** $\tilde{\zeta}$.

Proposition 15.20 (a) If V is a real form of $\tilde{V}$, then V is a real vector space with the dimension of V over R equal to the dimension of $\tilde{V}$ over C.

(b) If $\tilde{\zeta} : \tilde{g} \to gl(\tilde{V})$ is an irreducible representation of $\tilde{g}$, then any real form of $\tilde{\zeta}$, say $\zeta : g \to gl(V)$, is an irreducible representation of g.

(c) If $\tilde{\zeta}_1 : \tilde{g} \to gl(\tilde{V}_1)$ and $\tilde{\zeta}_2 : \tilde{g} \to gl(\tilde{V}_2)$ are two equivalent irreducible representations of $\tilde{g}$, then $\tilde{\zeta}_1$ is real for a real form g of $\tilde{g}$ if and only if $\tilde{\zeta}_2$ is real for g.

PROOF (a) and (b) are trivial to prove.

(c) Choose a $T : \tilde{V}_1 \to \tilde{V}_2$ such that $\tilde{\zeta}_2(X) = T\tilde{\zeta}_1(X)T^{-1}$ for all $X \in \tilde{g}$ and let $\zeta_1 : g \to gl(V_1)$ be a real form of $\tilde{\zeta}_1$. Then $V_2 = \{T(x) : x \in V_1\}$ is a real form of $\tilde{V}_2$ and for all $X \in g$, $x \in V_1$ we have $\tilde{\zeta}_2(X)T(x) = T\tilde{\zeta}_1(X)T^{-1}(T(x)) = T\tilde{\zeta}_1(X)(x) \in V_2$ as required.

Examples (5) Obviously the adjoint representation of any complex Lie algebra $\tilde{g}$ is real for all real forms of $\tilde{g}$.

(6) The representations of the complex Lie algebras of type A_n, B_n, C_n, and D_n discussed in example (3) above are real for their split forms because we know that $sl(n + 1, R)$, $\mathscr{B}(n, R)$, $\mathscr{C}(n, R)$, and $\mathscr{D}(n, R)$ are representations of the split real forms in terms of real matrices of the required type.

Exercises (5) Let $\tilde{g}$ be the complex Lie algebra of type A_n, g a compact real form of $\tilde{g}$, and $\tilde{\zeta} : \tilde{g} \to gl(n + 1, C)$ the usual representation of $\tilde{g}$ as the set of $(n + 1) \times (n + 1)$ complex matrices of trace 0. Show that $\tilde{\zeta}$ is not real for g.

(6) Show that both basic irreducible representations of a complex Lie algebra of type G_2 are real for both real forms.

3. Some Simple Real and Complex Lie Groups

In this section we make a few brief comments on how the results we have obtained for representations can be applied to Lie groups. We will concern ourselves only with simple Lie groups; that is, Lie groups whose Lie algebras are among the finite-dimensional simple Lie algebras we have discussed. Also see Definition 10.14.

Definition 15.21 A Lie group G is said to be a **linear Lie group** if there exists some finite-dimensional vector space V and some $G_1 \subset GL(V)$ with G and G_1 isomorphic as Lie groups.

Examples (1) If $\tilde{\zeta} : \tilde{g} \to gl(\tilde{V})$ is a representation of a simple complex Lie algebra $\tilde{g}$ and $\exp : gl(\tilde{V}) \to GL(\tilde{V})$ is the usual exponential map, then the subgroup of $GL(\tilde{V})$ generated by $\exp(\tilde{\zeta}(\tilde{g}))$ is clearly a complex linear Lie group. If $\zeta : g \to gl(V)$ is a real form of $\tilde{\zeta}$, then the subgroup of $GL(V)$ generated by $\exp(\zeta(g))$ is a real linear Lie group.

(2) For $n \geq 2$ let $\tilde{g}$ be a complex Lie algebra of type A_n and let $\tilde{\zeta}_1$, $\tilde{\zeta}_2 : \tilde{g} \to gl(n+1, C)$ be the two representations of example (4), Section 15.2. Thus $\tilde{\zeta}_1$ and $\tilde{\zeta}_2$ are not equivalent. The groups generated by $\exp(\tilde{\zeta}_1(\tilde{g}))$ and $\exp(\tilde{\zeta}_2(\tilde{g}))$ are equal to $SL(n+1, C)$. Therefore the complex linear Lie groups corresponding to the two inequivalent representations can be viewed as the same set of matrices.

REMARK (1) Recall that the centers of connected simple Lie groups are discrete and are finite if, in addition, the groups are compact. If G is a connected simple Lie group, G_1 is the simply connected covering group of G and Z is the center of G_1, then there exists a subgroup H of Z such that G is Lie isomorphic to G_1/H. If we are given a simply connected simple Lie group G_1 and ζ a representation of the Lie algebra g of G_1, then G, the group generated by $\exp(\zeta(g))$, is Lie isomorphic to some G_1/H. One reason that representations of Lie algebras are so important is that the subgroups H may not be isomorphic for inequivalent representations ζ. It is also known that for certain subgroups H of Z, the group G_1/H is not Lie isomorphic to the group generated by $\exp(\zeta(g))$ for any irreducible representation ζ of g. For certain subgroups H the groups G_1/H may even fail to be linear groups. It is a difficult problem determining the relationship between the real forms of irreducible representations of a simple Lie algebra and the corresponding subgroups of the center of the appropriate simply connected group. We will be satisfied stating a few results for the case that the Lie algebra is simple complex or is a compact real form. A very interesting set of tables containing the results for all real forms can be found in the book by Tits [1967].

Proposition 15.22 If g is a compact real form of a simple complex Lie algebra $\tilde{g}$ and $\zeta : g \to gl(V)$ is the real form of some irreducible representation of $\tilde{g}$, then G, the group generated by $\exp(\zeta(g))$, is a real compact Lie group; that is, G is compact as a manifold.

PROOF Since any covering group of a simple compact Lie group is compact (Section 12.6), we need only consider the case when ζ is the adjoint representation of g. Thus let G be generated by $\exp(\text{ad}(g))$ and recall that

$$\text{Kill}(T(X), T(Y)) = \text{Kill}(X, Y)$$

for all $X, Y \in g$ and $T \in G$. We can define a positive definite bilinear form on g by $B(X, Y) = -\text{Kill}(X, Y)$, and we have $G \subset O(g) = \{T \in GL(g) : B(T(X), T(Y)) = B(X, Y)\}$. Thus G must be compact since $O(g)$ is compact.

REMARKS (2) A much stronger result is known and is not too difficult to prove, namely a simple connected Lie group is compact if and only if its Lie algebra is a compact real form of some complex Lie algebra.

(3) We conclude this section by stating a theorem which lists the centers of simply connected, compact, simple Lie groups. Several texts, including those by Wolf [1967], Loos [1969], Freudenthal and de Vries [1969], and Tits [1967] give a proof of the theorem or parts of it. The last book mentioned also lists similar information for noncompact groups and lists the centers of the groups generated by $\exp(\zeta(V))$ for various representations of simple Lie algebras.

Theorem 15.23 Let G_1 be a connected, simply connected, simple, complex Lie group. Let G be the maximal compact subgroup of G_1, and let g_1 and g be the Lie algebras of G_1 and G, respectively. Then g is a compact real form of g_1 and G is simply connected. If $Z(G_1)$ and $Z(G)$ denote the centers of G_1 and G, then $Z(G_1) = Z(G)$ and these centers are listed in the Table 15.4.

TABLE 15.4

Type of g_1	$Z(G_1) = Z(G)$
A_n, $n \geq 1$	Z_{n+1}
B_n, $n \geq 2$	Z_2
C_n, $n \geq 3$	Z_2
D_n, n even, $n \geq 4$	$Z_2 \oplus Z_2$
D_n, n odd, $n \geq 5$	Z_4
E_6	Z_3
E_7	Z_2
E_8, F_4, G_2	$\{1\}$

APPENDIX

In such texts as those by Helgason [1962], Kobayashi and Nomizu [1963], Loos [1969], and Wolf [1967], a differential geometric approach to Lie groups is used. In this chapter we briefly introduce some of the basic concepts of differential geometry and relate them to some of the algebra we have developed. We give few proofs but sufficient references.

Thus, after generalizing differentiation in Euclidean space to "differentiation" on a manifold M by using a connection ∇, we discuss the basic concepts of geodesics, parallel translation, pseudo-Riemannian structures, and holonomy groups. Then we apply these results to the G-invariant connections on $M = G/H$ which is a reductive homogeneous space. The G-invariant connections are in one-to-one correspondence with certain nonassociative algebras which are rather general, as shown in Fig. A.1, where the dashed arrow denotes a local correspondence. Next we show how these algebras correspond to certain multiplications $\mu : G/H \times G/H \to G/H$ analogous to the Lie group results in Sections 1.6 and 5.3. These results generalize facts on Lie groups and Lie algebras and we indicate how simple nonassociative algebras are related to irreducible G-invariant connections on G/H and how to construct general results from these components. Finally we give a way of computing those nonassociative algebras which induce pseudo-Riemannian G-invariant connections on G/H in terms of a Jordan algebra of endomorphisms.

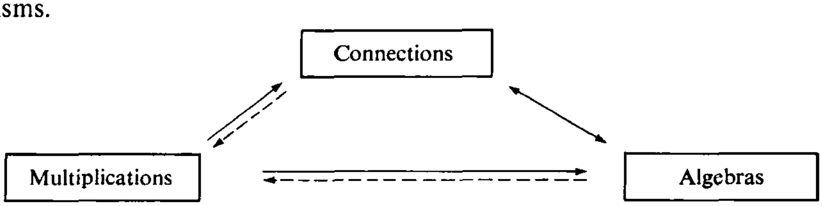


Fig. A. 1

1. Connections on Manifolds

We now consider the concept which generalizes differentiation in R^n to differentiation on a C^∞-manifold M. Then the geometric concepts of geodesics, curvature, torsion, holonomy, and Riemannian connections are discussed.

Definition A.1 Let M be a C^∞-manifold and let $D(M)$ be the vector space of C^∞-vector fields on M. An **affine connection** on M is an R-bilinear map $\nabla : D(M) \times D(M) \to D(M) : (X, Y) \to \nabla_X Y$ satisfying

$$\nabla_{fX+gY}(Z) = f\,\nabla_X Z + g\,\nabla_Y Z,$$
$$\nabla_X(fY) = f\,\nabla_X Y + (Xf)Y,$$

where $f, g \in C^\infty(M)$. The operator ∇_X is called **covariant differentiation** relative to X.

Note that for $M = R^n$, a vector field Y can be regarded as a function $Y: R^n \to R^n$; thus the above definition reduces to that given in Chapter 1. Also it is shown by Kobayashi and Nomizu [1963] that connections exist on a manifold.

In terms of a coordinate neighborhood U of $p \in M$ with coordinate function $x = (x_1, \ldots, x_n)$, the connection ∇ is determined by n^3 real-valued C^∞-functions Γ_{ij}^k on U by

$$\nabla_{\partial/\partial x_i}(\partial/\partial x_j) = \sum_k \Gamma_{ij}^k\, \partial/\partial x_k$$

where $\partial/\partial x_r$ are the usual coordinate vector fields on U.

These functions Γ_{ij}^k completely determine ∇ on U, for given vector fields $X = \sum a_i\, \partial/\partial x_i$ and $Y = \sum b_j\, \partial/\partial x_j$ on U, we can use the properties of ∇ to show

$$\nabla_X Y = \sum_k \sum_i [a_i(\partial b_k/\partial x_i + \sum_j \Gamma_{ij}^k b_j)]\, \partial/\partial x_k. \qquad (*)$$

Conversely, given any real-valued C^∞-functions Γ_{ij}^k on U, we can define $\nabla_X Y$ by (*). This can then be extended to all of M provided certain compatibility conditions are satisfied [Helgason, 1962, p. 27].

Definition A.2 Let M be a C^∞-manifold with affine connection ∇, let $\sigma : I \to M$ be a C^∞-curve in M with tangent vector field X; that is, $X(t) = \dot{\sigma}(t)$ for all t in the open interval I, and let J be a closed subinterval of I. A C^∞-vector field Y on σ is **parallel along** σ (restricted to J) if $(\nabla_X Y)(\sigma(t)) = 0$ for all $t \in J$. The curve σ is a **geodesic** if $(\nabla_X X)(\sigma(t)) = 0$ for all $t \in J$.

The proofs of the following results on parallel fields and geodesics can be found in the work of Helgason [1962], Hicks [1965] and Kobayashi and Nomizu [1963] and simply involve the solution to differential equations.

Theorem A.3 Let M be a C^∞-manifold with affine connection ∇, let $\sigma : I \to M$ be a C^∞-curve in M, and let $[a, b] \subset I$. For each $Y \in T(M, \sigma(a))$ there is a unique C^∞-vector field $Y(t)$ on $\sigma | [a, b]$ such that $Y = Y(a)$ and $Y(t)$ is parallel along σ (restricted to $[a, b]$). Furthermore the map

$$\sigma(a, b) : T(M, \sigma(a)) \to T(M, \sigma(b)) : Y(a) \to Y(b)$$

is a vector space isomorphism called **parallel translation** along σ from $\sigma(a)$ to $\sigma(b)$.

Outline of proof Without loss of generality we can assume σ lies in a coordinate neighborhood U and has no double points. For $x_1, \ldots, x_n$ coordinate functions on U, we can write $\sigma(t) = \sum \sigma_i(t)\, \partial/\partial x_i(\sigma(t))$ and $X(t) = \dot{\sigma}(t) = \sum \sigma_i'(t)\, \partial/\partial x_i(\sigma(t))$ and $Y(t) = \sum a_i(t)\, \partial/\partial x_i(\sigma(t))$. Then $Y(t)$ parallel along σ implies

$$0 = \nabla_X Y = \sum_k \Big(\sum_i \sigma_i'\, \partial a_k/\partial x_i + \sum_{i,j} \sigma_i' a_j \Gamma_{ij}^k\Big)\, \partial/\partial x_k$$

on σ. This gives for $t \in [a, b]$ the equation for the a_k,

$$da_k/dt + \sum_{i,j} a_j \sigma_i' \Gamma_{ij}^k = 0$$

for $k = 1, \ldots, n$ where we use the chain rule on $a_k(t) = a_k(\sigma_1(t), \ldots, \sigma_n(t))$. The unique solution to this system of linear differential equations gives the results concerning the isomorphism $\sigma(a, b)$. If σ is to be a geodesic, we obtain for $X(t) = Y(t)$ the system

$$d^2\sigma_k/dt^2 + \sum_{i,j} \Gamma_{ij}^k\, d\sigma_i/dt\, d\sigma_j/dt = 0$$

whose unique solution with specified initial conditions gives the following result.

Theorem A.4 Let M be a C^∞-manifold with affine connection ∇, let $p \in M$, and let $X \in T(M, p)$. Then for any real number a there exists a real number $\varepsilon > 0$ and a unique geodesic $\sigma : [a - \varepsilon, a + \varepsilon] \to M$ such that $\sigma(a) = p$ and $\dot{\sigma}(a) = X$.

Note that if we let $\sigma = \sigma(t, p, X, a)$ be the curve given by the above theorem, then from the theory of differential equations σ is a C^∞-function of the parameters t, p, X, and a. Also note that for $M = R^n$ and $\Gamma_{ij}^k = 0$ on M

we obtain from the above differential equations that the geodesics are straight lines. We now consider some functions related to affine connections.

Definition A.5 Let M be a C^∞-manifold with affine connection ∇ and let X, Y, Z be C^∞-vector fields on M. The **torsion tensor,** Tor is given by $\operatorname{Tor}(X,Y) = \nabla_X Y - \nabla_Y X - [X, Y]$ and the **curvature tensor,** R is given by

$$R(X, Y)Z = (\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X, Y]})Z.$$

We also use the notation $\operatorname{Tor}_\nabla$ and R_∇.

Note that $\operatorname{Tor}(X, Y)$ and $R(X, Y)Z$ are again vector fields which are multilinear in X, Y, Z and satisfy several algebraic identities. The following results can be found in the work of Hicks [1965] and Kobayashi and Nomizu [1963].

Theorem A.6 Let M be a C^∞-manifold.

(a) If $\bar{\nabla}$ is an affine connection, then there exists a unique connection ∇ with the same geodesics as $\bar{\nabla}$ and $\operatorname{Tor}_\nabla = 0$ on M.

(b) Two connections ∇ and $\bar{\nabla}$ on M are equal if and only if they have the same geodesics and $\operatorname{Tor}_\nabla = \operatorname{Tor}_{\bar{\nabla}}$ on M.

Definition A.7 Let M be a C^∞-manifold, let $p \in M$, and let $\Xi(p)$ denote the set of nondegenerate symmetric bilinear forms on $T(M, p)$. A **pseudo-Riemannian structure** on M is a map

$$\langle\, , \rangle : M \to \bigcup_{p \in M} \Xi(p)$$

such that for all $p, q \in M$, the bilinear forms $\langle\, , \rangle(p)$ and $\langle\, , \rangle(q)$ have the same index and such that $\langle\, , \rangle$ is C^∞ as follows. For each pair of C^∞-vector fields X, Y on M, the function $\langle X, Y\rangle : M \to R$ given by

$$\langle X, Y\rangle(p) = \langle X(p), Y(p)\rangle(p)$$

is C^∞ on M. We often use the notation $\langle\, , \rangle_p$ for the bilinear form $\langle\, , \rangle(p)$ on $T(M, p)$. If each bilinear form $\langle\, , \rangle(p)$ is positive definite, then $\langle\, , \rangle$ is a **Riemannian structure** on M. A **pseudo-Riemannian manifold** is a C^∞-manifold with a pseudo-Riemannian structure.

REMARKS (1) If the manifold M is connected, then we see that the index is automatically the same on each tangent space. This uses the C^∞-nature of the function $\langle\, , \rangle$ and consequently if the index did change at a point, then the form becomes degenerate at that point.

(2) If M is a paracompact C^∞-manifold, then there is a Riemannian structure on M; see the work of Singer and Thorpe [1967].

Example (1) Let G be a connected Lie group and H a closed (Lie) subgroup. Then on the homogeneous space G/H the diffeomorphisms $\tau(a) : G/H \to G/H : pH \to apH$ for $a \in G$ induce the tangent maps $T(\tau(a))(\bar{p}) : T(G/H, \bar{p}) \to T(G/H, \overline{ap})$. In particular, for $u \in H$ we use the notation $\tilde{\tau}(u) = T(\tau(u))(\bar{e}) : T(G/H, \bar{e}) \to T(G/H, \bar{e})$ and note $\tilde{\tau}(H) = \{\tilde{\tau}(u) : u \in H\}$ is a subgroup of $GL(T(G/H, \bar{e}))$. Let C be a nondegenerate symmetric bilinear form on $T(G/H, e)$ such that for all $X, Y \in T(G/H, \bar{e})$ we have $C(\tilde{\tau}(u)X, \tilde{\tau}(u)Y) = C(X, Y)$ and for $\bar{p} = \tau(a)\bar{e} \in G/H$ define a nondegenerate symmetric bilinear form $\langle\ ,\ \rangle(\bar{p})$ on $T(G/H, \bar{p})$ by

$$\langle U, V\rangle(\bar{p}) = C(T\tau(a^{-1})(\bar{p})U, T\tau(a^{-1})(\bar{p})V)$$

for $U, V \in T(G/H, \bar{p})$. Using $C(\tilde{\tau}(u)X, \tilde{\tau}(u)Y) = C(X, Y)$ for $X, Y \in T(G/H, \bar{e})$ we see that $\langle\ ,\ \rangle(\bar{p})$ is independent of the choice of $a \in G$ so that $\tau(a)\bar{e} = \bar{p}$. A homogeneous space with a pseudo-Riemannian metric $\langle\ ,\ \rangle$ given as above is called a **pseudo-Riemannian homogeneous space**. Note that when H is compact there exists a positive definite form C. Thus in this case G/H is a Riemannian homogeneous space. For another example, let g and h be semisimple and write $g = m + h$ with $m = h^{\perp}$ relative to the Killing form Kill of g. Then identifying m with $T(G/H, \bar{e})$, note that $C = \text{Kill}|m \times m$ is a nondegenerate symmetric bilinear form such that $C(\tilde{\tau}(u)X, \tilde{\tau}(u)Y) = C(X, Y)$ for all $X, Y \in m$. Thus we obtain a pseudo-Riemannian metric $\langle\ ,\ \rangle$ on G/H.

The proof of the following can be found in the work of Helgason [1962], Hicks [1965], and Kobayashi and Nomizu [1963].

Theorem A.8 Let M be a C^∞-pseudo-Riemannian manifold. Then there exists a unique affine connection ∇ satisfying the following conditions.

(a) The torsion Tor is zero.

(b) Parallel translation preserves the bilinear form on the tangent spaces; that is, if X, Y are parallel vector fields along a curve σ, then the function $\langle X, Y\rangle$ is a constant on σ.

The connection ∇ given above is called the **pseudo-Riemannian connection** relative to the pseudo-Riemannian structure on M. It should be noted that conditions (a) and (b) can be expressed as:

(a′) $\nabla_X Y - \nabla_Y X - [X, Y] = 0$;
(b′) $Z\langle X, Y\rangle = \langle\nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$

for C^∞-vector fields X, Y, Z on M.

We shall now show how parallel translation induces a group acting on a tangent space $T(M, p)$; a detailed expository discussion of this can be found in the text by Nomizu [1961]. Thus let ∇ be an affine connection on M and for $p \in M$ let $H(p)$ be the set of vector space isomorphisms of $T(M, p)$ obtained by parallel translation around all broken C^∞-curves σ which start and end at p. We shall let the same letter σ also denote the parallel translation along the closed curve σ. For two such parallel translations σ, τ along closed curves σ, τ at p, the parallel translation along the composite curve $\sigma\tau$ is just the endomorphism product of the parallel translations σ and τ. With this we see that $H(p)$ becomes a group of endomorphisms of $T(M, p)$ called the **holonomy group** of ∇ at p. The **restricted holonomy group** $H^0(p)$ is the subgroup of the holonomy group $H(p)$ obtained by restricting the parallel translations to closed curves which are homotopic to 0.

REMARKS (3) If M is connected, then for each $p, q \in M$ we have $H(p)$ is isomorphic to $H(q)$ as groups. In this case we define up to isomorphism the **holonomy group** of M, denoted by $\mathrm{Hol}(\nabla)$, by $\mathrm{Hol}(\nabla) = H(p)$ for some $p \in M$.

(4) The holonomy group $H(p)$ is a Lie group.

(5) Let M be a pseudo-Riemannian manifold with pseudo-Riemannian structure $\langle\, ,\rangle$ and with the corresponding pseudo-Riemannian connection ∇. Then we define the **holonomy group of the pseudo-Riemannian connection** at p to be those endomorphisms of $H(p)$ which are $\langle\, ,\rangle(p)$-orthogonal endomorphisms of $T(M, p)$. Thus in this case the elements of the holonomy group satisfy the additional condition of preserving the pseudo-Riemannian structure.

A discussion of the above remarks and the following results is given by Nomizu [1961] and Kobayashi and Nomizu [1963]. Let M be a connected Riemannian manifold with Riemannian structure $\langle\, ,\rangle$ and let $p, q \in M$. The **distance** $d(p, q)$ between the two points p and q is defined to be the infimum of the lengths of all broken C^1-curves joining p and q. [The length of a C^1-curve $\sigma(t)$ for $a \le t \le b$ is $\int_a^b < \dot\sigma(t), \dot\sigma(t) > dt$ where $\dot\sigma(t)$ denote the tangent vectors of the curve.] The function d satisfies the axioms of a metric and gives the same topology on M as the original manifold topology.

Definition A.9 A Riemannian manifold M is **complete** if the above metric d is complete; that is, if every Cauchy sequence relative to d has a limit point.

REMARKS (6) Every compact Riemannian manifold is complete and also every Riemannian homogeneous space G/H is complete.

(7) If M is a connected complete Riemannian manifold, then any two points $p, q \in M$ can be joined by a geodesic whose length equals $d(p, q)$.

Completeness is used in the global version of the following decomposition theorem due to deRham:

Definition A.10 Let the manifold M have affine connection ∇ and let $H(p)$ be the holonomy group at $p \in M$. Then $T(M, p)$ is **holonomy irreducible** if the group $H(p)$ acts irreducibly on $T(M, p)$. If M is connected, then M is **holonomy irreducible** if $H(p)$ is holonomy irreducible relative to some reference point $p \in M$.

Theorem A.11 Let M be a connected Riemannian manifold with Riemannian connection ∇.

(a) There is an orthogonal direct sum decomposition $T(M, p) = T_0 + T_1 + \cdots + T_s$, where $T_0 = \{X \in T(M, p) : AX = X \text{ all } A \in H(p)\}$ and for $1 \le i \le s$ each T_i is $H(p)$-invariant and irreducible.

(b) If M is simply connected, then the decomposition in (a) is unique up to order and $H(p)$ is the direct product $H_0 \times H_1 \times \cdots \times H_s$ of normal subgroups, where H_0 is the identity on $T(M, p)$ and each H_i for $1 \le i \le s$ acts trivially on T_j for $j \ne i$ and irreducibly on T_i.

(c) If M is a simply connected and complete Riemannian manifold, then M is isometric to the direct product $M_0 \times M_1 \times \cdots \times M_s$ where M_0 is a Euclidean space (possibly of dimension 0) and the M_i for $1 \le i \le s$ are simply connected complete holonomy irreducible Riemannian manifolds. This decomposition is unique up to order.

2. Connections on Homogeneous Spaces and Nonassociative Algebras

Let G be a connected Lie group with Lie algebra g and H a closed (Lie) subgroup with Lie subalgebra h. Then the pair (G, H) or (g, h) is called a **reductive pair** if there exists a subspace m of g so that $g = m + h$ (subspace direct sum) and $(\operatorname{Ad} H)(m) \subset m$. This last condition gives the condition $[hm] \subset m$ in terms of Lie algebras. The corresponding manifold $M = G/H$ is called a **reductive homogeneous space**. The main example we shall be using is when H is semisimple and connected. In particular, when G is also semisimple we can decompose $g = m + h$ where $m = h^{\perp}$ relative to the Killing form of g.

In order to consider G-invariant connections on G/H we use the following definition [Helgason, 1962; Kobayashi and Nomizu, 1963].

Definition A.12 Let $\phi : M \to M$ be a diffeomorphism of the manifold M with connection ∇. Then ϕ is an **affine map** or a **connection preserving map** if $\phi'(\nabla_X Y) = \nabla_{\phi' X} \phi' Y$ for all $X, Y \in D(M)$, where ϕ' is given by $\phi' X = [(T\phi)X] \circ \phi^{-1}$; see Section 2.7.

Definition A.13 Let G/H be a reductive homogeneous space and for $a \in G$ let $\tau(a) : G/H \to G/H : xH \to axH$. A connection ∇ on G/H is a **G-invariant connection** if for all $a \in G$ the functions $\tau(a)$ are affine maps.

Let (A, α) denote a nonassociative algebra A where α is the bilinear multiplication on the underlying vector space A. Then continuing the same notation, the G-invariant connections on G/H are given by the following result and we use the notation of Nomizu [1954]; also note the work of Kobayashi and Nomizu [1968].

Theorem A.14 Let G/H be a reductive homogeneous space with a fixed Lie algebra decomposition $g = m + h$ such that $(\mathrm{Ad}\, H)(m) \subset m$ and let $\tilde{\mathrm{A}}\mathrm{d}\, H$ denote the induced maps on m. Then there exists a one-to-one correspondence between the set of all G-invariant affine connections on G/H and the set of all nonassociative algebras (m, α) with $\tilde{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$ which is the group of automorphisms of the algebra (m, α).

REMARKS (1) To obtain $\alpha(X, Y)$ for $X, Y \in m$ we evaluate $\nabla_{X^*} Y^*$ at $\bar{e} = eH$ for certain vector fields X^*, Y^* determined by X, Y on a neighborhood N^* of $\bar{e}$.

(2) Let H consist of the identity so that $G = G/H$ is an n-dimensional Lie group with Lie algebra g. Then any nonassociative algebra structure on the n-dimensional vector space g yields a G-invariant connection on G and conversely. Thus as with Lie groups and Lie algebras, we can use the results on nonassociative algebras to study G-invariant connections.

(3) As discussed in example (4), Section 7.3, the subspace m in the decomposition $g = m + h$ can be made into an anticommutative algebra as follows. For $X, Y \in m$ let $[XY] = XY + h(X, Y)$, where $XY = [XY]_m$ (respectively $h(X, Y) = [XY]_h$) is the projection of $[XY]$ in g into m (respectively h). This algebra is related to connections by the following result [Nomizu, 1954; Kobayashi and Nomizu, 1968].

Theorem A.15 On a reductive homogeneous space G/H with fixed decomposition $g = m + h$, there exists one and only one G-invariant connection which has zero torsion tensor and such that the curves $\gamma(t) = \exp tX$ for all $X \in m$ project by $\pi : G \to G/H$ into geodesics $\gamma^*(t) = \pi\gamma(t)$ in G/H. In this case we have $\alpha(X, Y) = \frac{1}{2}XY$. This connection is called the **canonical connection of the first kind** and we shall denote the corresponding algebra by $(m, \frac{1}{2}XY)$.

These algebras are not too difficult to compute. Thus continuing exercise (7), Section 12.2 and example (1), Section 6.4 we have for $g = so(n)$, $h = so(p)$ the decomposition $g = m + h$ with $m = h^{\perp}$ consisting of matrices of the form

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & 0 \end{bmatrix}$$

and h consisting of matrices of the form

$$\begin{bmatrix} 0 & 0 \\ 0 & B_{22} \end{bmatrix}.$$

Consequently $XY = [X, Y]_m$ is of the form

$$\begin{bmatrix} [X_{11}, Y_{11}] + [X_{12}, Y_{21}] & X_{11}Y_{12} - Y_{11}X_{12} \\ X_{21}Y_{11} \;-\; Y_{21}X_{11} & 0 \end{bmatrix}.$$

REMARK (4) For a G-invariant connection determined by the algebra (m, α) the torsion and curvature tensors evaluated at $\bar{e} = eH$ are given by

$$\begin{aligned} \mathrm{Tor}(X, Y) &= \alpha(X, Y) - \alpha(Y, X) - XY, \\ R(X, Y)Z &= \alpha(X, \alpha(Y, Z)) - \alpha(Y, \alpha(X, Z)) - \alpha(XY, Z) - [h(X, Y)\, Z] \end{aligned}$$

where $X, Y, Z \in m$ and XY, $h(X, Y)$ are given above. Letting $R(X, Y)$ be the endomorphism given by $R(X, Y) : m \to m : Z \to R(X, Y)Z$ as above and letting $L(X) : m \to m : Y \to \alpha(X, Y)$ we have the following result [Nomizu, 1961; Kobayashi and Nomizu, 1968].

Theorem A.16 The Lie algebra of the holonomy group $\mathrm{Hol}(\nabla)$ of a G-invariant connection ∇ on G/H determined by the algebra (m, α) is the smallest Lie algebra h^* of endomorphisms of m such that:

(a) $R(X, Y) \in h^*$ for all $X, Y \in m$;
(b) $[L(X), h^*] \subset h^*$ for all $X \in m$.

Corollary A.17 If ∇ is a holonomy irreducible connection of the first kind given by the algebra $(m, \frac{1}{2}XY)$, then this algebra is simple or a zero algebra.

PROOF Briefly, suppose $XY \not\equiv 0$ and let $k = L(m) + \mathrm{Der}(m)$ where $L(m)$ is the Lie transformation algebra of m; see Section 7.2. Then k is a Lie algebra of endomorphisms of m such that $[L(X), k] \subset k$ for all $X \in m$ and such that

$$R(X, Y) = [L(X), L(Y)] - L(XY) - D(h(X, Y))$$

is in k for all $X, Y \in m$. This uses the formula for $R(X, Y)Z$ and part (vi) of Exercise (3), Section 7.3, which shows $D(h(X, Y)) = \mathrm{ad}\, h(X, Y)|_m \in \mathrm{Der}(m)$. Thus from the above theorem, $h^* \subset k$. Now suppose m is not simple. Then from the work of Sagle and Winter [1967] there exists a proper ideal n of m which is $\mathrm{Der}(m)$-invariant. Thus $h^*n \subset kn \subset n$ which contradicts the holonomy irreducibility.

We now consider pseudo-Riemannian reductive homogeneous spaces G/H with the pseudo-Riemannian structure given in example (1), Section 16.1 and have the following result [Nomizu, 1954; Kobayashi and Nomizu, 1968].

Theorem A.18 Let G/H be a reductive homogeneous space with fixed Lie algebra decomposition $g = m + h$. Let C be a symmetric nondegenerate bilinear form on m which gives the pseudo-Riemannian metric $\langle\ ,\ \rangle$ on G/H and let ∇ be the corresponding pseudo-Riemannian connection. Then the algebra (m, α) induced by ∇ is given by

$$\alpha(X, Y) = \tfrac{1}{2}XY + U(X, Y)$$

where $U(X, Y) = U(Y, X)$ is uniquely determined. Furthermore α satisfies

$$C\big(\alpha(Z, X), Y\big) + C\big(X, \alpha(Z, Y)\big) = 0 \quad \text{and} \quad C\big((\bar{\mathrm{A}}\mathrm{d}\, u)X, (\bar{\mathrm{A}}\mathrm{d}\, u)Y\big) = C(X, Y)$$

for all $X, Y, Z \in m$ and $u \in H$. Denote such an algebra by (m, α, C).

Remarks (5) From the work of Nomizu [1954] and Kobayashi and Nomizu [1968], $U(X, Y)$ is given by

$$2C\big(U(X, Y), Z\big) = C(ZX, Y) + C(X, ZY)$$

for $X, Y, Z \in m$.

(6) For a pseudo-Riemannian connection of the first kind the algebra (m, α) satisfies $\alpha(X, Y) = \tfrac{1}{2}XY$, and for $X, Y, Z \in m$, $U \in h$,

$$C(XZ, Y) = C(X, ZY) \qquad \text{and} \qquad C\big((\mathrm{ad}\, U)X, Y\big) = -C\big(X, (\mathrm{ad}\, U)Y\big).$$

Note this gives the possibility of applying Theorem 12.10 concerning associative forms. Examples of such algebras occur when g and h are semisimple, $m = h^{\perp}$ relative to the Killing form of g, and $C = \mathrm{Kill}\,|\, m \times m$.

(7) Let $(m, \tfrac{1}{2}XY)$ be an algebra with positive definite form C satisfying the above conditions. Let s be an ideal of m so that $ss = \{0\}$ and s is maximal relative to this property. Then write $m = s + m'$ an orthogonal direct sum and note that since C is an associative form, m' is an ideal of m such that $sm' = \{0\}$. Now $C\,|\, m' \times m'$ is nondegenerate and m' contains no ideals n with $nn = \{0\}$. Thus we can apply Theorem 12.10 to conclude $m' = m_1 + \cdots + m_k$ where each m_i is a simple ideal. However, this corresponds to the decomposition of Theorem A.11 and we have the following result; note the work of Kobayashi and Nomizu [1968].

Corollary A.19 Let $M = G/H$ be a simply connected Riemannian reductive homogeneous space with the connection of the first kind induced by the algebra $(m, \tfrac{1}{2}XY)$ and let the Lie algebra of the holonomy group be nonzero.

(a) Then M is isometric to the direct product $M_0 \times M_1 \times \cdots \times M_k$ of Riemannian reductive homogeneous spaces such that the holonomy group acts trivially on M_0 and irreducibly on M_i for $1 \leq i \leq k$.

(b) Let $m = m_0 + m_1 + \cdots + m_k$ be the corresponding decomposition of $m \cong T(M, \bar{e})$ and let $s = m_0 + m_{i_1} + \cdots + m_{i_t}$ for suitable corresponding $M_0, M_{i_1}, \ldots, M_{i_t}$. Then the algebra $m = s + m'$ where $sm = \{0\}$ and $m' = m_{j_1} + \cdots + m_{j_p}$ where the ideals $m_{j_1}, \ldots, m_{j_p}$ are simple algebras corresponding to the remaining holonomy irreducible spaces $M_{j_1}, \ldots, M_{j_p}$.

3. Multiplicative Systems and Connections

We now show how a multiplication on a reductive homogeneous space G/H yields a G-invariant connection by obtaining an algebra (m, α) from the multiplication. This is analogous to the way the Lie algebra arises from the multiplication in a Lie group. Let G/H be a reductive homogeneous space and let

$$\mu : G/H \times G/H \to G/H$$

be an analytic function such that $\mu(\bar{e}, \bar{e}) = \bar{e}$ where $\bar{e} = eH$. Then the structure $(G/H, \mu)$ is called a **multiplication**. If $\tau(H) = \{\tau(u) : u \in H\}$ is contained in $\mathrm{Aut}(G/H, \mu)$, the automorphism group of the multiplication, then$(G/H, \mu)$ is called a **multiplicative system**. Analogous to the computations in Sections 1.6 and 5.3, we show how a multiplicative system yields an algebra (m, α) with $\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$ and consequently yields a G-invariant connection.

Thus let $\pi : G \to G/H$ be the natural projection and let $g = m + h$ be a fixed decomposition. Then from Section 6.4 we know that for the map $\psi = \exp | m$ there exists an open neighborhood D of 0 in m which is mapped homeomorphically into G under ψ and such that π maps $\psi(D)$ homeomorphically onto an open neighborhood N^* of $\bar{e}$ in G/H. Consequently from the analyticity of μ and $\pi \circ \psi$ there exists a neighborhood U of 0 in m such that for all $X, Y \in U$

$$\mu(\pi \exp X, \pi \exp Y) = \pi \exp F(X, Y)$$

is in N^* where $F : U \times U \to D$ is analytic at $\theta = (0, 0)$ in $m \times m$.

From this μ is determined locally by F which has the Taylor's series expansion

$$F(X, Y) = F(\theta) + F^1(\theta)(X, Y) + \tfrac{1}{2}F^2(\theta)(X, Y)^{(2)} + \cdots$$

for X, $Y \in U$. As in Section 1.6 we use $\mu(\bar{e}, \bar{e}) = \bar{e}$ to see $F(\theta) = 0$ and we also have

$$F^1(\theta)(X, Y) = PX + QY$$

where $PX = F^1(\theta)(X, 0)$, $QY = F^1(\theta)(0, Y)$ and

$$F^2(\theta)(X, Y)^{(2)} = F^2(\theta)(X, 0)^{(2)} + F^2(\theta)(0, Y)^{(2)} + 2F^2(\theta)[(X, 0), (0, Y)].$$

As before we define for X, $Y \in m$,

$$\alpha(X, Y) = F^2(\theta)[(X, 0), (0, Y)]$$

and see this is a bilinear function $\alpha : m \times m \to m$. Therefore the multiplication μ on G/H determines a nonassociative algebra (m, α).

Note that the converse is true locally. Thus given a nonassociative algebra (m, α) we let $F(X, Y) = X + Y + \alpha(X, Y)$. Then there is a neighborhood U of 0 in m so that for X, $Y \in U$, $\bar{\mu}(\pi \exp X, \pi \exp Y) = \pi \exp F(X, Y)$ defines an analytic local multiplication $\bar{\mu}$ on some neighborhood N^* of $\bar{e}$. Note that $2F^2(\theta)[(X, 0), (0, Y)] = F^2(\theta)(X, Y)^{(2)} = 2\alpha(X, Y)$ in this case.

To obtain a connection from the algebra (m, α) induced by the multiplicative system $(G/H, \mu)$, we need $\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$ which follows from $\tau(H) \subset \mathrm{Aut}(G/H, \mu)$ as follows. First for $x \in G$, $u \in H$ note

$$\tau(u)\pi(x) = uxH = uxu^{-1}(uH) = \pi\phi(u)(x)$$

where $\phi(u) : G \to G : x \to uxu^{-1}$ is the inner automorphism of the group G determined by u. Also recall from Section 7.3 that $\mathrm{Ad}\, u = T(\phi(u))(e)$ and $\phi(u)(\exp X) = \exp(\mathrm{Ad}\, u(X))$. Next we have for X, Y near enough 0 in m

$$\begin{aligned}\tau(u)\mu(\pi \exp X, \pi \exp Y) &= \mu(\tau(u)\pi \exp X, \tau(u)\pi \exp Y)\\ &= \mu(\pi\phi(u) \exp X, \pi\phi(u) \exp Y)\\ &= \mu(\pi \exp \mathrm{Ad}\, uX, \pi \exp \mathrm{Ad}\, uY)\\ &= \pi \exp F(\mathrm{Ad}\, u(X), \mathrm{Ad}\, u(Y))\end{aligned}$$

and also

$$\begin{aligned}\tau(u)\mu(\pi \exp X, \pi \exp Y) &= \tau(u)\pi \exp F(X, Y)\\ &= \pi\phi(u) \exp F(X, Y)\\ &= \pi \exp \mathrm{Ad}\, u(F(X, Y)).\end{aligned}$$

Thus we conclude $\mathrm{Ad}\, u(F(X, Y)) = F(\mathrm{Ad}\, u(X), \mathrm{Ad}\, u(Y))$ using $(\mathrm{Ad}\, h)(m) \subset m$. This implies $\bar{\mathrm{A}}\mathrm{d}\, u(\alpha(X, Y)) = \alpha(\bar{\mathrm{A}}\mathrm{d}\, u(X), \bar{\mathrm{A}}\mathrm{d}\, u(Y))$ using the definition of $\alpha(X, Y) = F^2(\theta)[(X, 0), (0, Y)]$.

Conversely, if $\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$, then for $F(X, Y) = X + Y + \alpha(X, Y)$ we see $\mathrm{Ad}\, u(F(X, Y)) = F(\mathrm{Ad}\, uX, \mathrm{Ad}\, uY)$ and consequently $\tau(u)$ is an automorphism of the previously defined local multiplication $\bar{\mu}$. Thus we obtain a local multiplicative system corresponding to the algebra (m, α) with

$\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$. We summarize the results as follows, and for more details see the work of Sagle and Schumi [to appear].

Theorem A.20 Let $(G/H, \mu)$ be a multiplicative system and let $g = m + h$ be the fixed Lie algebra decomposition. Let μ be given locally by $\mu(\pi \exp X, \pi \exp Y) = \pi \exp F(X, Y)$ where $F(X, Y) = PX + QY + \frac{1}{2}F^2(\theta)(X, Y)^{(2)} + \cdots$ and let $\alpha(X, Y) = F^2(\theta)[(X, 0), (0, Y)]$.

(a) Then α is a bilinear function which determines an algebra (m, α) so that $\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$. Thus the multiplicative system $(G/H, \mu)$ induces a G-invariant connection on G/H.

(b) Conversely, an algebra (m, α) with $\bar{\mathrm{A}}\mathrm{d}\, H \subset \mathrm{Aut}(m, \alpha)$ determines a local multiplicative system so that when the multiplication $\bar{\mu}$ is represented by $\bar{\mu}(\pi \exp X, \pi \exp Y) = \pi \exp F(X, Y)$ we obtain $\alpha(X, Y) = F^2(\theta)[(X, 0), (0, Y)]$.

REMARKS (1) If the multiplicative system is a Lie group G with μ the group multiplication, then $F(X, Y)$ is given by the Campbell–Hausdorf formula and $\alpha(X, Y) = \frac{1}{2}[XY]$. Thus the geodisics relative to the induced connection are of the form $\gamma(t) = \exp tX$ for $X \in g$.

(2) Conditions on the algebra (m, α) can be obtained by considering identities on the multiplicative system $(G/H, \mu)$. Thus $(G/H, \mu)$ is **power-associative** if $\bar{e}$ is an identity element and each $\bar{x} \in G/H$ generates an associative subsystem containing $\bar{e}$; that is, $(G/H, \mu)$ is a power associative "H-space." In particular, this means the powers $\bar{x}^n$, for n a positive integer, are uniquely defined. Thus for each $\bar{x} \in G/H$ the map $n \to \bar{x}^n$ is a well-defined homomorphism of the positive integers under addition into $(G/H, \mu)$. In this case it can be shown that the algebra (m, α) is anticommutative. Thus the geodesics relative to the corresponding connection are of the form $\pi \exp tX$ for $X \in m$; note the work of Nomizu [1954, Section 10].

A specific nonassociative example arises from the Cayley division algebra over R by considering the set M of elements of norm 1 (where the positive definite norm is used). In a paper by Schumi and Walde [to appear] it is shown that $M = S^7 = G/H$ where G is of type B_3 and H is of type G_2, and the multiplication μ is the nonassociative multiplication of the Cayley algebra. Also $(G/H, \mu)$ is a multiplicative system and the corresponding algebra (m, α) is given by $\alpha(X, Y) = \frac{1}{2}XY$ where this anticommutative multiplication satisfies $J(X, Y, XZ) = J(X, Y, Z)X$ where $J(X, Y, Z) = (XY)Z + (YZ)X + (ZX)Y$. An anticommutative algebra satisfying this identity is called a "Malcev algebra"; note the work by Sagle [1965].

(3) Since many properties of a Lie group are given by its Lie algebra considered as G-invariant vector fields, we can also use the functions $l(X) : G/H \to T(G/H)$ analogous to that discussed in exercise (3), Section 2.7,

exercise (1), Section 2.8, and exercise (1), Section 5.1 for the multiplicative system $(G/H, \mu)$. Thus for $\bar{p} \in G/H$ and $X \in m$ we have $l(X)(\bar{p}) = [(T\mu)(\bar{p}, \bar{e})](0, X)$. Note as in the exercises $l(X)$ is a vector field if and only if $\bar{e}$ is a right identity element; that is, $\mu(\bar{x}, \bar{e}) = \bar{x}$ all $\bar{x} \in G/H$. Many properties of $l(X)$ are given in the work of Sagle and Schumi [to appear]. In particular, if (G, μ) is a multiplicative system with every $l(X)$ a G-invariant vector field, then the corresponding algebra (g, α) is given by $\alpha(X, Y) = \frac{1}{2}[XQY]$ where $QY = F^1(\theta)(0, Y)$. If α is not the zero function and the corresponding connection is holonomy irreducible, then Q is nonsingular.

4. Riemannian Connections and Jordan Algebras

In the previous sections we showed how to find G-invariant connections on G/H in terms of multiplications and nonassociative algebras. For the pseudo-Riemannian case these algebras can be explicitly computed in terms of the algebra $(m, \frac{1}{2}XY)$ and a certain Jordan algebra as follows; see the work of Sagle [to appear]. Let $(m, \frac{1}{2}XY, B)$ denote the algebra which induced the pseudo-Riemannian connection of the first kind on G/H according to Theorem A.18 where B is used instead of the C in Theorem A.18. Now suppose G/H has another pseudo-Riemannian connection given by the algebra (m, α, C). Then since the forms are nondegenerate there exists a unique $S \in GL(m)$ such that for all $X, Y \in m$

$$C(X, Y) = B(SX, Y).$$

Furthermore using the symmetry and Ad H-invariance of B and C we have for all $u \in H$,

$$S^b = S \qquad \text{and} \qquad [\bar{\mathrm{A}}\mathrm{d}\, u, S] = 0 \tag{$*$}$$

where b denotes the adjoint relative to B and $\bar{\mathrm{A}}\mathrm{d}\, u = \mathrm{Ad}\, u|m$. Using remark (5), Section A.2 we obtain the formula

$$2\alpha(X, Y) = XY + S^{-1}[X(SY) - (SX)Y].$$

Conversely given $S \in GL(m)$ satisfying $(*)$ we can define C and α as above to obtain the algebra (m, α, C) which induces a pseudo-Riemannian G-invariant connection on G/H.

Now the set

$$J = \{T \in \mathrm{End}(m) : T^b = T \text{ and } [\bar{\mathrm{A}}\mathrm{d}\, u, T] = 0 \text{ all } u \in H\}$$

forms a Jordan algebra of endomorphisms relative to the usual multiplication $S_1 \cdot S_2 = \frac{1}{2}(S_1 S_2 + S_2 S_1)$; see Section 14.2 concerning type F_4.

REMARKS (1) Using the fact that $S = \exp T$ is invertible for $T \in J$ we obtain a correspondence between the above connections and elements of J. Also note for $S = I$ we obtain the original connection determined by $(m, \frac{1}{2}XY, B)$.

(2) In case $XY \equiv 0$; that is, G/H is a symmetric space, the above results still hold for J.

Using Lemma 7.15 and the obvious variation of Definition 9.12, the results of Sagle [to appear] give the following:

Theorem A.21 Let (G, H) be a reductive pair with decomposition $g = m + h$ such that $\tilde{\mathrm{A}}\mathrm{d}\, H$ is completely reducible in m. Then J is a semisimple Jordan algebra; that is, a direct sum of simple Jordan algebras.

Corollary A.22 Let (G, H) be a reductive pair with G and H semisimple. Let the decomposition $g = m + h$ be given by $m = h^{\perp}$ relative to the Killing form Kill of g and let $B = \mathrm{Kill}|m \times m$. Then the algebra $(m, \frac{1}{2}XY, B)$ determines a pseudo-Riemannian connection of the first kind and the Jordan algebra J is semisimple.

Example (1) Let $G = SO(n)$ and let $H = SO(p)$ for $p < n - 1$. Then $g = so(n)$ the $n \times n$ skew-symmetric matrices and embed $h = so(p)$ as in example (1), Section 6.4. Thus $m = h^{\perp}$ is the set of matrices of the form

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & 0 \end{bmatrix}$$

and the spaces K_0 and K_2 of matrices of the form

$$\begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & X_{12} \\ X_{21} & 0 \end{bmatrix},$$

respectively, are J-invariant. Thus $J = J_0 \oplus J_2$ where J_0 and J_2 are isomorphic to the simple Jordan algebra of symmetric $r \times r$ matrices where $r = n - p$. As far as we know, it is an open problem to classify the Jordan algebras J for the reductive pairs (g, h) with g and h semisimple.

REFERENCES

Auslander, L. and MacKenzie, R. (1963). "Introduction to Differentiable Manifolds." McGraw-Hill, New York.

Barns, W. (1965). "Introduction to Abstract Algebra." Heath, Indianapolis, Indiana.

Bishop, R. and Goldberg, S. (1968). "Tensor Analysis on Manifolds." Macmillan, New York.

Bourbaki, N. (1960). "Groupes et Algèbres de Lie," Chapter 1. Hermann, Paris.

Bourbaki, N. (1968). "Groupes et Algèbres de Lie," Chapters 4–6. Hermann, Paris.

Bourbaki, N. (1972). "Groupes et Algèbres de Lie," Chapters 2–3. Hermann, Paris.

Brickell, F. and Clark, R. (1970). "Differential Manifolds, An Introduction." Van Nostrand-Reinhold, Princeton, New Jersey.

Chevalley, C. (1946). "Theory of Lie Groups." Princeton Univ. Press, Princeton, New Jersey.

Cohn, P. (1957). "Lie Groups." Cambridge Univ. Press, London and New York.

Dieudonné, J. (1960). "Foundations of Modern Analysis." Academic Press, New York.

Eichler, M. (1968). A new proof of the Baker–Campbell–Hausdorff formula. *J. Math. Soc. Japan* **20**, 23–25.

Freudenthal, H. and De Vries, H. (1969). "Linear Lie Groups". Academic Press, New York.

Hausner, M. and Schwartz, J. (1968). "Lie Groups and Lie Algebras." Gordon & Breach, New York.

Helgason, S. (1962). "Differential Geometry and Symmetric Spaces." Academic Press, New York.

Hicks, N. (1965). "Notes on Differential Geometry." Van Nostrand-Reinhold, Princeton, New Jersey.

Hewitt, E. and Ross, K. (1963). "Abstract Harmonic Analysis," Volume I. Springer-Verlag, Berlin and New York.

Hochschild, G. (1965). "The Structure of Lie Groups." Holden-Day, San Francisco, California.

Husain, T. (1966). "Introduction to Topological Groups." Saunders, Philadelphia, Pennsylvania.

Jacobson, N. (1953). "Lectures in Abstract Algebra," Volumes I, II. Van Nostrand-Reinhold, Princeton, New Jersey.

Jacobson, N. (1962). "Lie Algebras." Wiley, New York.

Jacobson, N. (1971a). "Exceptional Lie Algebras." Dekker, New York.

Jacobson, N. (1971b). "Structure and Representations of Jordan Algebras." Amer. Math. Soc., Providence, Rhode Island.
Kobayashi, S. and Nomizu, K. (1963). "Foundations of Differential Geometry." Volume I. Wiley, New York.
Kobayashi, S. and Nomizu, K. (1968). "Foundations of Differential Geometry," Volume II. Wiley, New York.
Lang, S. (1962). "Introduction to Differentiable Manifolds." Wiley, New York.
Lang, S. (1965). "Algebra." Addison-Wesley, Reading, Massachusetts.
Lang, S. (1968). "Analysis I." Addison-Wesley, Reading, Massachusetts.
Loos, O. (1969). "Symmetric Spaces." Benjamin, New York.
MacLane, S. (1963). "Homology." Springer-Verlag, Berlin and New York.
McCarty, G. (1967). "Topology, An Introduction with Applications to Topological Groups." McGraw-Hill, New York.
Montgomery, D. and Zippin, L. (1955). "Topological Transformation Groups." Wiley, New York.
Nomizu, K. (1954). Invariant affine connections on homogeneous spaces. *Amer. J. Math.*, **76**, 33–65.
Nomizu, K. (1961). Recent developments in the theory of connections and holonomy groups. *Advances in Math.*, Academic Press, New York.
Paley, H. and Weichsel, P. (1966). "A First Course in Abstract Algebra." Holt, New York.
Pontryagin, L. (1946). "Topological Groups." Princeton Univ. Press, Princeton, New Jersey.
Sagle, A. (1965). On simple extended Lie Algebras over fields of characteristic zero. *Pacific J. Math.* **2**.
Sagle, A. Jordan algebras and connections on homogeneous spaces, to appear.
Sagle, A. and Schumi, J. Multiplications on homogeneous spaces, nonassociative algebras and connections. *Pacific J. Math.* **45**.
Sagle, A. and Winter, D. (1967). On homogeneous spaces and reductive subalgebras of simple Lie algebras. *Trans. Amer. Math. Soc.*, **128**, 142–147.
Samelson, H. (1969). "Notes on Lie Algebras." Van Nostrand-Reinhold, Princeton, New Jersey.
Schafer, R. D. (1966). "An Introduction to Nonassociative Algebras." Academic Press, New York.
Schumi, J. and Walde, R. Some homogeneous spaces related to Cayley algebras, to appear.
Seminaire Sophus Lie (1955). "Théorie des Algèbres de Lie, Topologie des Groupes de Lie." Ecole Normale Supérieure, Paris.
Serre, J-P. (1965). "Lie Algebras and Lie Groups." Benjamin, New York.
Serre, J-P. (1966). "Algèbres de Lie Semi-simples Complexes." Benjamin, New York.
Singer, I. and Thorpe, J. (1967). "Lecture Notes on Elementary Topology and Geometry." Scott, Foresman, Glenview, Illinois.
Spivak, M. (1965). "Calculus on Manifolds." Benjamin, New York.
Tits, J. (1965). "Liesche Gruppen und Algebren." Univ. Bonn, Bonn.
Tits, J. (1967). "Tabellen zu den einfachen Lie gruppen und ihren Darstellungen." Springer-Verlag, Berlin and New York.
Tondeur, Ph. (1965). "Introduction to Lie Groups and Transformation Groups." Springer-Verlag, Berlin and New York.
Warner, F. (1971). "Foundations of Differentiable Manifolds and Lie Groups." Scott, Foresman, Glenview, Illinois.
Winter, D. (1972). "Abstract Lie Algebras." M.I.T. Press, Cambridge, Massachusetts.
Wolf, J. (1967). "Spaces of Constant Curvature." McGraw-Hill, New York.