Pathways in Mathematics

Urs Frauenfelder
Otto van Koert

The Restricted Three-Body Problem and Holomorphic Curves

Birkhäuser

Pathways in Mathematics

Birkhäuser

Each "Pathways in Mathematics" book offers a roadmap to a currently well developing mathematical research field and is a first-hand information and inspiration for further study, aimed both at students and researchers. It is written in an educational style, i.e., in a way that is accessible for advanced undergraduate and graduate students. It also serves as an introduction to and survey of the field for researchers who want to be quickly informed about the state of the art. The point of departure is typically a bachelor/masters level background, from which the reader is expeditiously guided to the frontiers. This is achieved by focusing on ideas and concepts underlying the development of the subject while keeping technicalities to a minimum. Each volume contains an extensive annotated bibliography as well as a discussion of open problems and future research directions as recommendations for starting new projects

More information about this series at http://www.springer.com/series/15133

Urs Frauenfelder • Otto van Koert

The Restricted Three-Body Problem and Holomorphic Curves

Birkhäuser

Urs Frauenfelder
Mathematical Institute
University Augsburg
Augsburg, Germany

Otto van Koert
Department of Mathematics
Seoul National University
Gwanak-gu, Republic of Korea

ISSN 2367-3451 ISSN 2367-346X (electronic)
Pathways in Mathematics
ISBN 978-3-319-72277-1 ISBN 978-3-319-72278-8 (eBook)
https://doi.org/10.1007/978-3-319-72278-8

Library of Congress Control Number: 2018953209

Mathematics Subject Classification (2010): 53D42, 32Q65, 70F07

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Danpung, Sujin, Yukichi and Yuko

Contents

Chapter 1

Introduction

1.1 The Birkhoff conjecture

The study of the restricted three-body problem has a long history. Nevertheless many questions still remain to be solved. We quote from page 328 of Birkhoff's inspiring essay. [39]

> "This state of affairs seems to me to make it probable that the restricted problem of three bodies admit of reduction to the transformation of a discoid into itself as long as there is a closed oval of zero velocity about J(upiter), and that in consequence there always exists at least one direct periodic orbit of simple type."

Translated into modern language, Birkhoff asks if below the first critical energy value in each bounded component of the restricted three-body problem there exists a disk-like global surface of section. The Poincaré return map associated to the disk-like global surface of section then has by Brouwer's translation theorem at least one fixed point. These fixed points give rise to direct periodic orbits.

What do we know a century after Birkhoff published his essay about his question? Unfortunately, we still cannot answer Birkhoff's question with hundred percent certainty in the affirmative. Nevertheless, the modern methods of symplectic geometry make it quite likely that the answer to this century old question will be found in the near future. The purpose of these notes is to make young ambitious researchers familiar with the main players of Birkhoff's question, namely the restricted three-body problem and disk-like global surfaces of section, and to introduce them to the modern techniques of symplectic geometry, which reduce Birkhoff's question to questions about symplectic embeddings and systolic geometry.

Through the introduction of holomorphic curves into symplectic topology Gromov revolutionized the subject [105]. Hofer [120] and Hofer, Wysocki, and Zehnder [122, 124, 128] later extended the theory of holomorphic curves to symplectizations of contact manifolds and discovered in [126] that disk-like global

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_1

surfaces of section can be constructed with the help of holomorphic curves. The theory of Hofer, Wysocki, and Zehnder was later refined by Siefring with the discovery of his intersection number and by Hryniewicz with the introduction of fast finite energy planes. These theories allow us to find sufficient and necessary conditions for a periodic orbit to bound a disk-like global surface of section. In [9] the moon was “contacted”, i.e., it was shown that for energies below and slightly above the first critical value the bounded components of the energy hypersurface of the regularized restricted three-body problem are of contact type. This result enables us to apply holomorphic curve techniques to the restricted three-body problem.

The restricted three-body problem studies the dynamics of a massless body attracted by two massive bodies according to Newton’s law of gravitation. For example, the “massless body” can be imagined as a satellite and the two massive bodies are the earth and the moon. Another option is to think of the “massless body” as the moon and the two massive bodies are the sun and the earth. Or one could think of the “massless body” as the planet Tatooine attracted by the two stars Tatoo I and Tatoo II as in the Star Wars saga. In fact, when the planet Tatooine in the Star Wars saga first appeared, nobody knew if such planets exist in reality. Due to amazing progress in the search for exoplanets in the last decades we now know that such worlds with two suns exist, see for example [74].

1.2 The power of holomorphic curves

The question about the existence of a global surface of section is a question about all orbits not just periodic ones. However, holomorphic curves seem to confirm Poincaré’s fantastic insight [205] that periodic orbits in some sense are the “skeleton” of the dynamics. The power of the technology of holomorphic curves lies in the fact that they reduce the question about existence of global surfaces of section to questions about periodic orbits.

Not every periodic orbit can bound a disk-like global surface of section. Indeed, if a periodic orbit bounds a disk-like global surface of section, it necessarily has to satisfy some topological conditions. The mere fact that it is the boundary of an embedded disk implies that it is unknotted. Another obstruction is that it necessarily has to be linked to every other periodic orbit. A more subtle obstruction which comes from contact topology is that its self-linking number has to be -1. Now suppose that the dynamics can be interpreted as the Reeb flow of a star-shaped hypersurface $\Sigma \subset \mathbb{C}^2$, which we know to hold true in the case of interest for Birkhoff’s conjecture [9]. Then results of Hryniewicz and Salomão [135, 136] show that these necessary conditions are basically sufficient as well.

The idea behind this result, which goes back to the ground breaking work of Hofer, Wysocki, and Zehnder [126], is the following. One considers finite energy planes in the symplectization $\Sigma \times \mathbb{R}$. These are finite energy solutions of a holomorphic curve equation for maps from the plane to the symplectization. Note that Σ has three dimensions, so the dimension of $\Sigma \times \mathbb{R}$ is four, while holomorphic planes are two-dimensional. A pair of two-dimensional objects in a four-dimensional space generically intersects in isolated points. For holomorphic curves the local intersection index is positive. On the other hand, to define an intersection number for holomorphic planes that is invariant under homotopies is not an easy task, because a plane is noncompact and one has to worry about intersections at infinity. In his celebrated approach to the Weinstein conjecture [120] Hofer already noted that the finiteness of energy for a plane guarantees that asymptotically the projection of a finite energy plane to Σ converges to a periodic Reeb orbit. If one fixes the asymptotics, Siefring [220] managed to define an intersection number for finite energy planes which has all the properties one would expect for an intersection number from the closed case. In particular, it is homotopy invariant and moreover, in view of positivity of intersection for holomorphic curves, its vanishing guarantees that two curves with different image do not intersect at all. Even more is true. If Siefring's intersection number for two finite energy planes vanishes, then even their projections to Σ do not intersect unless the images coincide.

One now fixes a periodic Reeb orbit and considers the moduli space of finite energy planes asymptotic to this fixed periodic orbit. There are various symmetries on this moduli space. First there are the reparametrizations of maps from the plane to $\Sigma \times \mathbb{R}$ and moreover, there is an $\mathbb{R}$-action which comes from the obvious $\mathbb{R}$-action on the second factor of the target. If one considers the moduli space modulo these symmetries, the dimension of the space becomes two less than the Conley–Zehnder index of the asymptotic Reeb orbit. In particular, if the Conley–Zehnder index is three, the dimension of the moduli space is one. Moreover, due to some automatic transversality miracle it can be shown in this case that the moduli space is actually a manifold, so that it is a disjoint union of circles and intervals. In case the Conley–Zehnder index is bigger than three, Hryniewicz [133] found a way to associate a subspace to the space of finite energy planes, which after dividing out the symmetries becomes a one-dimensional manifold. Namely, he discovered the notion of "fast finite energy" planes; these are finite energy planes which have a fast exponential asymptotic decay, where the decay rate is chosen in such a way to guarantee that the projection of the finite energy plane to Σ is an immersion transverse to the Reeb vector field.

Now one considers the moduli space modulo symmetries of fast finite energy planes to a fixed asymptotic Reeb orbit γ with vanishing Siefring self-intersection number. By the properties of Siefring's intersection number the projections of these fast finite energy planes to Σ do not intersect and therefore build a local foliation of Σ. To get a global foliation one has to understand the compactness properties of this moduli space. A compactification of the moduli space of finite energy planes is provided by the SFT-compactness theorem [43]. In the case of fast finite energy planes with vanishing Siefring self-intersection number, SFT-compactness implies that if the moduli space is noncompact, a sequence of fast finite energy planes has to converge to a negatively punctured fast finite energy plane and at these punctures the projection of the punctured fast finite energy plane converges to periodic Reeb orbits, which are unlinked to the Reeb orbit γ. In particular, if γ is linked to every other periodic Reeb orbit this scenario cannot occur and therefore the moduli space of fast finite energy planes has to be compact. In case it is not empty, we get an open book decomposition of Σ whose binding is the periodic Reeb orbit γ. Each page of this open book corresponds to the projection of a fast finite energy plane to Σ. It turns out that each page is a global surface of section for the Reeb flow.

1.3 Systolic inequalities and symplectic embeddings

As we have seen in the previous section in order to prove compactness of the moduli space of fast finite energy planes with vanishing Siefring self-intersection number asymptotic to the Reeb orbit γ, one needs to make sure that γ is linked to every other periodic Reeb orbit. However, looking at the SFT-compactness theorem more closely reveals that if the moduli space is noncompact and therefore it contains a negatively punctured fast finite energy plane in its closure, then the periodic orbits at the negative punctures satisfy additional conditions

(i) The periods of the periodic orbits at the negative punctures are less than the period of γ.

(ii) At least one periodic orbit at a negative puncture has Conley–Zehnder index less than or equal to 2.

In particular, we deduce from (i) that if γ has the shortest period among all periodic orbits, compactness is guaranteed. The periodic orbit of smallest period is referred to as the *systole*. Hence the Birkhoff conjecture prompts the following question, which if answered positively, actually would imply the Birkhoff conjecture with the help of holomorphic curve techniques.

Question: *Does the retrograde periodic orbit represent the systole?*

We point out that we only expect a positive answer to the question for energy values below the first critical value. Indeed, above the first critical value a periodic orbit bifurcates out of the critical point known as Lyapunov orbit, which at least for energies slightly above the first critical value represents the systole.

In Riemannian geometry the systole is just the shortest geodesic and there is a huge literature on the subject, see for example [37]. The introduction of systolic questions to the contact world is rather new and started with the work of Álvarez Paiva and Balacheff [13]. How much the Riemannian world and the contact world can differ was recently realized through the discoveries of Abbondandolo, Bramham, Hryniewicz, and Salomão [2]. The work of Lee [158] represents the first steps to approach the systole in Hill's lunar problem via Symplectic homology.

In view of (ii) a different approach to the Birkhoff conjecture is to rule out periodic orbits of Conley–Zehnder index 2. The following definition is due to Hofer, Wysocki, and Zehnder [126]

Definition 1.3.1. A starshaped hypersurface $\Sigma \subset \mathbb{C}^2$ is called *dynamically convex* if the Conley–Zehnder indices of all periodic Reeb orbits are at least 3.

Armed with this notion we can now ask

Question: *Is the restricted three-body problem dynamically convex?*

Again this question only applies to energies below the first critical value, despite the fact that the dynamics of the restricted three-body problem just above the first critical value can be interpreted as the Reeb flow of a hypersurface, [9]. The culprit is again the Lyapunov orbit, which bifurcates out of the critical point. This periodic orbit has Conley–Zehnder index 2!

Only in rare cases dynamical convexity can be checked directly. Indeed, first finding all periodic orbits and then compute their Conley–Zehnder indices is in general not doable. However, Hofer, Wysocki, and Zehnder [126] found a much more tractable criterion.

Theorem [Hofer, Wysocki, Zehnder]: *A strictly convex hypersurface $\Sigma \subset \mathbb{C}^2$ is dynamically convex.*

It is interesting to note that while dynamical convexity is a symplectic notion, the concept of convexity is not. Indeed, it can well happen that a non-convex starshaped hypersurface admits a different symplectic embedding which is convex. On the other hand, the theorem of Hofer, Wysocki, and Zehnder tells us that a starshaped hypersurface in the complex plane which has a periodic orbit of Conley–Zehnder index less than three does not admit a convex symplectic embedding at all. To our knowledge the following problem is still open.

Question: *Does there exist a dynamically convex starshaped hypersurface $\Sigma \subset \mathbb{C}^2$ which does not admit a convex symplectic embedding at all?*

The question about symplectic embeddings was one of the driving forces for the development of symplectic topology. Starting with Gromov's celebrated non-squeezing theorem [105], the question which symplectic manifolds can be symplectically embedded into each other is a highly active area of research, see for example [214, 215] for a survey.

Moreover, the question about convex symplectic embeddings and the systole are related. Indeed, it is known that in the convex case the Hofer–Zehnder capacity coincides with the systole [131], and obtaining estimates for the systole in the convex case is strongly related to the famous Viterbo conjecture [237, 199].

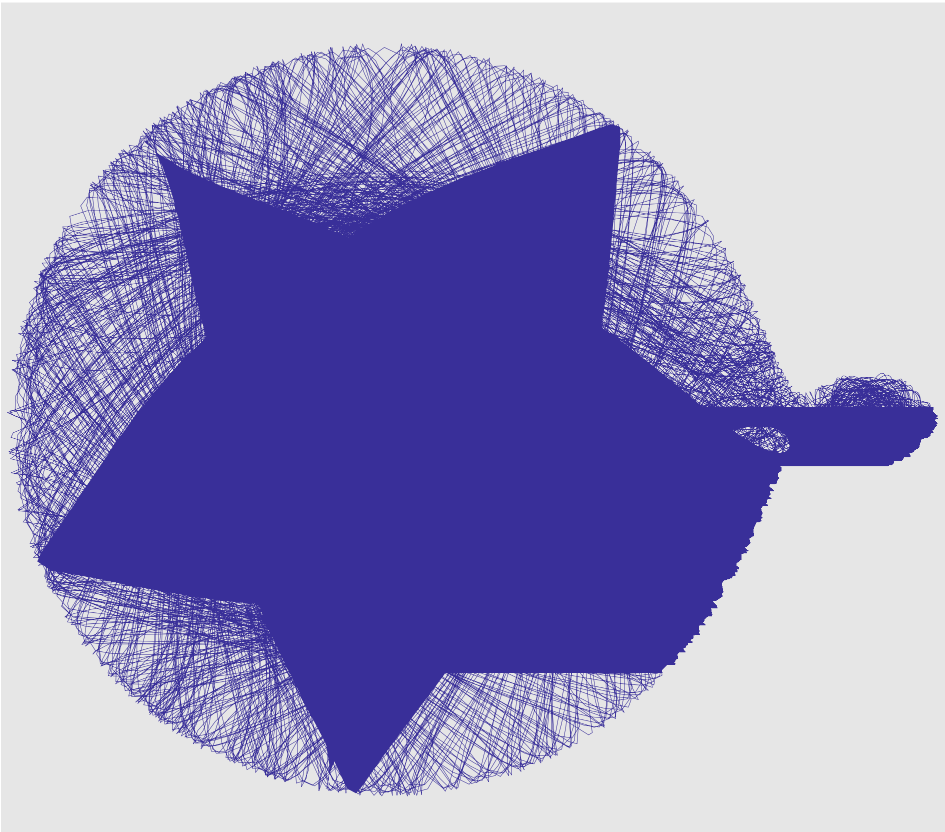
What do we know about dynamical convexity for the restricted three-body problem? For energies below the first critical value the Levi-Civita regularization provides us with an embedding into $\mathbb{C}^2$ and it was shown in [9] that the corresponding hypersurface is starshaped. In view of the theorem of Hofer, Wysocki, and Zehnder the natural question is if this embedding is maybe even convex. In fact this is sometimes true. It was shown in [6] that for small mass ratios around the small mass the embedding is convex and therefore dynamical convexity holds. On the other hand, this fails around the heavy body, even in the limit when the mass of the light body is zero. This limit is the rotating Kepler problem, namely the Kepler problem in rotating coordinates. In fact it was noted in [7] that the Levi-Civita embedding of the rotating Kepler problem is not convex. On the other hand in the same paper, the Conley–Zehnder indices of all periodic orbits were computed and the outcome was that the rotating Kepler problem is dynamically convex. The question arose if the rotating Kepler problem might provide an example of a dynamically convex starshaped hypersurface in $\mathbb{C}^2$ which does not admit a convex symplectic embedding. This turned out to be wrong as well. In fact, in [94] a convex embedding for the rotating Kepler problem was found. This embedding used a combination of the Ligon–Schaaf and Levi-Civita embedding.

1.4 Beyond the Birkhoff conjecture

Birkhoff's assumption that there is a closed oval of zero velocity around each of the massive bodies precludes a satellite from traveling from the earth to the moon. This assumption of Birkhoff is equivalent to the satellite having an energy below the first critical value. Although we essentially only study this case in this monograph, our ultimate goal is actually to go above the first critical value. The binding orbit of the global surface of section that Birkhoff had in mind is the retrograde periodic orbit. Now it is impossible that the retrograde periodic orbit binds a global surface of section above the first critical point. The reason is that from the first critical point a new periodic orbit bifurcates, which is known as the Lyapunov orbit. The retrograde periodic orbit and the Lyapunov orbit are unlinked, and therefore the retrograde orbit cannot bound a global surface of section anymore. Here is what we expect to happen. The holomorphic curves giving rise to global surfaces of section below the first critical value break precisely at the Lyapunov orbit and give rise to a finite energy foliation similar as in [130]. Interesting research in this direction was carried out in [84, 203]. If this picture is correct, then it follows from the theory developed in [130] that the stable and unstable manifold of the Lyapunov orbit intersect. For small mass ratios this phenomenon was first noted by Conley [60] and McGehee [174], and this fact turned out to have fantastic applications to space mission design.

The traditional way from the earth to the moon uses the Hohmann transfer [132]. This transfer, described in detail in [28], uses two engine impulses; one at the beginning to bring the spacecraft to the transfer orbit and one at the end to stop. This method only takes advantage of the dynamics of the two-body problem, which have been well known since the times of Kepler. The Hohmann transfer was for example successfully applied for the Apollo Moon landings. It is fast, but one of its drawbacks is that it uses a lot of fuel, not just at the beginning, but also at the end of the journey to stop. Especially the fuel that one needs to stop the rocket has to be carried during its whole journey, which is very expensive. There is also some danger. Namely, if stopping the rocket does not work out properly, it either crashes into the moon or just flies by. It was Conley [59, 60, 61] who first propagated the idea to use the dynamics of the restricted three-body problem to find low energy transit orbits to the moon. In this spirit, in 1978 the spacecraft ISEE-3 (International Sun-Earth Explorer-3) was brought to a halo orbit at a Lagrange point of the earth-sun system using three-body trajectories [80, 81]. It was Belbruno who found the first realistic low energy transit orbit from an orbit around the earth to an orbit around the moon [32, 33]. Although many people at first were skeptical about these new methods, the rescue of the Japanese Hiten mission [35] impressively proved that chaotic motion can be applied to real mission design. In January 1990 the Japanese launched their first lunar mission. There were two robotic spacecrafts involved in this mission: MUSES-A (later on renamed Hiten) and MUSES-B (later on renamed Hagoromo). MUSES-B was supposed to go to the moon, while MUSES-A was to stay on an orbit around the earth as a communication relay with MUSES-B. Unfortunately, contact with MUSES-B was lost and only MUSES-A remained. Since MUSES-A was never supposed to go to the moon, it had only little fuel, too little to travel to the moon on a standard route. Nevertheless, on October 2, 1991, MUSES-A, now with the name Hiten, successfully arrived at the moon after a long travel which took advantage of chaotic motion in both the earth-moon-satellite system and in the sun-earth-satellite system. The many ups and downs in this thrilling story can be nicely read in Belbruno's book *Fly me to the moon* [34]. While the traditional space mission design is referred to as *patched conics*, namely patched two-body problems whose solutions are given by conic sections, the new approach to space mission design pioneered by Belbruno can be thought of as *patched restricted three-body problems.* We refer the reader to the beautiful book by Koon, Lo, Marsden, and Ross [152] for more information on this new paradigm in space mission design.

Acknowledgement. Part of these notes are extracted from lectures the first named author held at the University of Augsburg in the winter term 2014/15 and the two authors held at the CIMPA-NIMS Research school on dynamical systems which took place in Daejeon 25.5.–3.6.2015. We would like to thank heartily all the participants of these lectures. You greatly helped that these notes took shape. We sincerely thank CIMPA and NIMS for making this research school possible. We immensely profited from discussions with many people, in particular, Peter Albers, Edward Belbruno, Alain Chenciner, Hagyun Choi, Kai Cieliebak, Etienne Ghys, Joel Fish, Helmut Hofer, Umberto Hryniewicz, Jungsoo Kang, Seongchan Kim, Junyoung Lee, Thomas Ligon, Pedro Salomão, Felix Schlenk, Martin Schwingenheuer, Richard Siefring, Francesco Topputo, Lei Zhao.



Chapter 2

Symplectic Geometry and Hamiltonian Mechanics

There are many excellent books on Symplectic geometry like for example [5, 19, 50, 88, 131, 161, 172, 240]. Here we will give an overview of the features that are essential for our purposes.

2.1 Symplectic manifolds

The archetypical example of a symplectic manifold is the cotangent bundle of a smooth manifold. Assume that N is a manifold, by physicists also referred to as the *configuration space*. The *phase space* is the cotangent bundle T^*N, and this space comes endowed with a canonical one-form $\lambda \in \Omega^1(T^*N)$, called the *Liouville one-form*. It is defined as follows. Abbreviate by $\pi\colon T^*N \to N$ the footpoint projection. If we take $e \in T^*N$ and $\xi \in T_eT^*N$, the tangent space of T^*N at e, then the differential of the footpoint projection at e is a linear map

$$d\pi(e)\colon T_eT^*N \to T_{\pi(e)}N.$$

Interpreting e as a vector in $T^*_{\pi(e)}N$, the dual space of $T_{\pi(e)}N$, we can pair it with the vector $d\pi(e)\xi \in T_{\pi(e)}N$ and define

$$\lambda_e(\xi) := e\big(d\pi(e)\xi\big).$$

In canonical coordinates $(q,p) = (q_1, \dots, q_n, p_1, \dots, p_n)$ of T^*N, where $n = \dim N$, the Liouville one-form becomes

$$\lambda(q,p) = \sum_{i=1}^n p_i dq_i.$$

The canonical symplectic form on T^*N is the exterior derivative of the Liouville one-form

$$\omega = d\lambda.$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_2

In canonical coordinates it has the form

$$\omega = \sum_{i=1}^{n} dp_i \wedge dq_i.$$

The symplectic form ω has the following properties. It is *closed*, i.e., $d\omega = 0$. This is immediate because $d\omega = d^2\lambda = 0$. Furthermore, it is *non-degenerate* in the sense that if $e \in T^*N$ and $\xi \neq 0 \in T_eT^*N$, then there exists $\eta \in T_eT^*N$ such that $\omega(\xi, \eta) \neq 0$. These two properties become the defining properties of a symplectic structure on a general manifold M, which does not need to be a cotangent bundle.

Definition 2.1.1. A *symplectic manifold* is a pair (M, ω) where M is a manifold and $\omega \in \Omega^2(M)$ is a two-form satisfying the following two conditions

(i) ω is closed.
(ii) ω is non-degenerate.

The two-form ω is called the *symplectic structure* on M.

The assumption that ω is non-degenerate immediately implies that a symplectic manifold is even-dimensional. In other words an odd-dimensional manifold never admits a symplectic structure.

2.2 Symplectomorphisms

Symplectic manifolds form a category whose morphisms are given by symplectomorphisms, which are defined as follows.

Definition 2.2.1. Assume that (M_1, ω_1) and (M_2, ω_2) are two symplectic manifolds. A *symplectomorphism* $\phi\colon M_1 \to M_2$ is a diffeomorphism satisfying $\phi^*\omega_2 = \omega_1$.

We discuss three examples of symplectomorphisms.

2.2.1 Physical transformations

Suppose that N_1 and N_2 are manifolds and $\phi\colon N_1 \to N_2$ is a diffeomorphism, for example a change of coordinates of the configuration space. If $x \in N_1$, then the differential

$$d\phi(x)\colon T_xN_1 \to T_{\phi(x)}N_2$$

is a vector space isomorphism. Dualizing we get the vector space isomorphism

$$d\phi(x)^*\colon T^*_{\phi(x)}N_2 \to T^*_xN_1.$$

We now define

$$d_*\phi\colon T^*N_1 \to T^*N_2$$

as follows. If $\pi_1 \colon T^*N_1 \to N_1$ is the footpoint projection and $e \in T^*N_1$, then

$$d_*\phi(e) := (d\phi(\pi_1(e))^*)^{-1}e. \tag{2.1}$$

If λ_1 is the Liouville one-form on T^*N_1 and λ_2 is the Liouville one-form on T^*N_2 one checks that

$$(d_*\phi)^*\lambda_2 = \lambda_1. \tag{2.2}$$

Because the exterior derivative commutes with pullback, we obtain

$$(d_*\phi)^*\omega_2 = (d_*\phi)^*d\lambda_2 = d(d_*\phi)^*\lambda_2 = d\lambda_1 = \omega_1$$

which shows that $d_*\phi$ is a symplectomorphism.

Equation (2.2) might be rephrased by saying that $d_*\phi$ is an *exact symplectomorphism*, i.e., a symplectomorphism which preserves the primitives of the symplectic forms. The notion of exact symplectomorphism in a general symplectic manifold does, however, not make sense, since usually the symplectic form ω has no primitive. In fact, if the symplectic manifold (M,ω) is closed, the non-degeneracy of the symplectic form ω implies that the closed form ω induces a non-vanishing class $[\omega] \in H^2_{dR}(M)$, the second de Rham cohomology group of M. In particular, ω cannot be exact. Indeed, suppose by contradiction that $\omega = d\lambda$ is exact. Then if the dimension of M is $2n$, we obtain using the fact that ω is closed and Stokes' theorem

$$\int_M \omega^n = \int_M \omega^{n-1} \wedge d\lambda = \int_M d(\omega^{n-1} \wedge \lambda) = 0,$$

which contradicts the fact that ω^n is a volume form on M. The latter fact follows from the non-degeneracy assumption on the symplectic form.

2.2.2 The switch map

The second example of a symplectomorphism is the switch map

$$\sigma \colon T^*\mathbb{R}^n \to T^*\mathbb{R}^n.$$

Namely, if we identify the cotangent bundle $T^*\mathbb{R}^n$ with $\mathbb{R}^{2n}$ with global coordinates $(q,p) = (q_1,\dots,q_n,p_1,\dots,p_n)$ and symplectic form $\omega = \sum_{i=1}^n dp_i \wedge dq_i$, then the switch map is given by

$$\sigma(q,p) = (-p,q)$$

which is actually a linear symplectomorphism on $\mathbb{R}^{2n}$. Note that the switch map is not a physical transformation. In physical terms the q variables, i.e., the variables on the configuration space, are referred to as the position variables, while the p variables are referred to as the momenta. Hence the switch map interchanges the roles of the momenta and the positions. We will see, that the switch map plays a major role in Moser's regularization of two-body collisions. Note that to define the switch map it is important to have global coordinates on the configuration space. There is no way to define the switch map on the cotangent bundle T^*N of a general manifold N.

2.2.3 Hamiltonian transformations

The third example of symplectomorphisms we discuss are Hamiltonian transformations. Suppose that (M, ω) is a symplectic manifold and $H \in C^\infty(M, \mathbb{R})$. Physicists call smooth functions on a symplectic manifold *Hamiltonians*. The interesting point about Hamiltonians is that we can associate to them a vector field $X_H \in \Gamma(TM)$ which is implicitly defined by the condition

$$dH = \omega(\cdot, X_H).$$

Note that the assumption that the symplectic form is non-degenerate guarantees that X_H is well defined. The vector field X_H is called *Hamiltonian vector field*. Let us assume for simplicity in the following that M is closed. Under this assumption the flow of the Hamiltonian vector field exists for all times, i.e., we get a smooth family of diffeomorphisms

$$\phi_H^t \colon M \to M, \quad t \in \mathbb{R}$$

defined by the conditions

$$\phi_H^0 = \mathrm{id}_M, \qquad \frac{d}{dt}\phi_H^t(x) = X_H(\phi_H^t(x)),\ t \in \mathbb{R},\ x \in M.$$

An important property of the Hamiltonian flow is that the Hamiltonian H is preserved under it. If one interprets the Hamiltonian as the energy then this means the energy is conserved.

Theorem 2.2.2 (Preservation of energy)**.** *For $x \in M$ it holds that $H(\phi^t(x))$ is constant, i.e., independent of $t \in \mathbb{R}$.*

Proof. Differentiating we obtain

$$\begin{aligned} \frac{d}{dt}H(\phi_H^t(x)) &= dH(\phi_H^t(x))\frac{d}{dt}\phi_H^t(x) \\ &= dH(\phi_H^t(x))X_H(\phi_H^t(x)) \\ &= \omega(X_H, X_H)(\phi_H^t(x)) \\ &= 0, \end{aligned}$$

where the last equality follows from antisymmetry of the two-form ω. □

The next theorem tells us that the diffeomorphisms ϕ_H^t are symplectomorphisms. The intuition from physics might be that a Hamiltonian system has *no friction*.

Theorem 2.2.3. *For every $t \in \mathbb{R}$ it holds that $(\phi_H^t)^*\omega = \omega$.*

Proof. Note that

$$\frac{d}{dt}(\phi_H^t)^*\omega = (\phi_H^t)^*\mathcal{L}_{X_H}\omega$$

where $\mathcal{L}_{X_H}\omega$ is the Lie derivative of the symplectic form with respect to the Hamiltonian vector field. Using Cartan's formula we obtain by taking advantage of the assumption that ω is closed and the definition of X_H

$$\mathcal{L}_{X_H}\omega = \iota_{X_H}d\omega + d\iota_{X_H}\omega = -d^2H = 0.$$

This proves the theorem. □

2.3 Examples of Hamiltonians

2.3.1 The free particle and the geodesic flow

Assume that (N, g) is a Riemannian manifold. Define

$$H_g\colon T^*N \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|_g^2$$

where $|\cdot|_g$ denotes the norm induced by the metric g on the cotangent bundle of N. In terms of physics this is just the *kinetic energy*.

The flow of this Hamiltonian corresponds to the geodesic flow of the metric g on N. To describe this relation we assume for simplicity that N is compact, in order to ensure that the flows exist for all times. If $q \in N$ and $v \in T_qN$ we denote by

$$q_v\colon \mathbb{R} \to N$$

the unique geodesic meeting the initial conditions

$$q_v(0) = q, \quad \partial_t q_v(0) = v.$$

The *geodesic flow*

$$\Psi_g^t\colon TN \to TN$$

is the map

$$(q,v) \mapsto (q_v(t), \partial_t q_v(t)).$$

The metric g gives rise to a bundle isomorphism

$$\Phi_g\colon TN \to T^*N, \quad (q,v) \mapsto (q, g_q(v,\cdot)).$$

This allows us to interpret the geodesic flow as a map from the cotangent bundle to itself; we define the (co)geodesic flow as

$$\phi_g^t := \Phi_g\Psi_g^t\Phi_g^{-1}\colon T^*N \to T^*N.$$

Theorem 2.3.1. *The Hamiltonian flow of the kinetic energy H_g coincides with the geodesic flow in the sense that $\phi^t_{H_g} = \phi^t_g$ for every $t \in \mathbb{R}$.*

Proof. Recall that if $q \in C^\infty([0,1], N)$ is a geodesic, then in local coordinates it satisfies the geodesic equation which is the second-order ODE

$$\partial_t^2 q^\ell + \Gamma^\ell_{ij}(q)\partial_t q^i \partial_t q^j = 0. \tag{2.3}$$

Here we use Einstein summation convention. Moreover, the Γ^ℓ_{ij} are the Christoffel symbols which are determined by the Riemannian metric by the formula

$$\Gamma^\ell_{ij} = \frac{1}{2} g^{\ell k}(g_{ki,j} + g_{jk,i} - g_{ij,k}). \tag{2.4}$$

If we interpret the geodesic equation as a first-order ODE on the tangent bundle of N, i.e., if we consider $(q, v) \in C^\infty([0,1], TN)$ with $q(t) \in N$ and $v(t) \in T_{q(t)}N$, then (2.3) becomes

$$\begin{cases} \partial_t v^\ell + \Gamma^\ell_{ij}(q) v^i v^j = 0 \\ \partial_t q = v. \end{cases} \tag{2.5}$$

We next rewrite (2.5) as an equation on the cotangent bundle instead of the tangent bundle. For this purpose we introduce

$$p_i = g_{ij} v^j.$$

Using the p's instead of the v's and the formula (2.4) for the Christoffel symbols equation (2.5) translates to

$$\begin{cases} \partial_t g^{\ell i} p_i + \frac{1}{2} g^{\ell k}(g_{ki,j} + g_{jk,i} - g_{ij,k}) g^{im} p_m g^{jn} p_n = 0 \\ \partial_t q^i = g^{ij} p_j. \end{cases} \tag{2.6}$$

In view of the identity

$$g^{ij} g_{j\ell} = \delta^i_\ell,$$

where δ^i_ℓ is the Kronecker Delta, we obtain the relation

$$g^{ij}_{,k} g_{j\ell} + g^{ij} g_{j\ell,k} = 0$$

from which we deduce

$$g^{ij}_{,k} = -g^{i\ell} g^{jm} g_{\ell m,k}.$$

Plugging this formula in the first equation in (2.6) we obtain

$$\begin{aligned} 0 &= g^{\ell i}_k \partial_t q^k p_i + g^{\ell i} \partial_t p_i - \frac{1}{2}\big(g^{\ell m}_{,j} g^{jn} + g^{\ell n}_{,i} g^{im} - g^{mn}_{,k} g^{k\ell}\big) p_m p_n \\ &= g^{\ell i}_{,k} g^{kj} p_j p_i + g^{\ell i} \partial_t p_i - \frac{1}{2} g^{\ell i}_{,k} g^{kj} p_j p_i \\ &= g^{\ell i} \partial_t p_i + \frac{1}{2} g^{\ell i}_{,k} g^{kj} p_j p_i \end{aligned}$$

from which we get

$$\begin{aligned}\partial_t p_r &= g_{r\ell} g^{\ell i} \partial_t p_i \\ &= -\frac{1}{2} g_{r\ell} g^{\ell k} g^{mn}_{,k} p_m p_n \\ &= -\frac{1}{2} \delta^k_r g^{mn}_{,k} p_m p_n \\ &= -\frac{1}{2} g^{mn}_{,r} p_m p_n.\end{aligned}$$

Therefore the geodesic equation (2.6) on the cotangent bundle becomes

$$\begin{cases} \partial_t p_i = -\frac{1}{2} g^{mn}_{,i} p_m p_n \\ \partial_t q^i = g^{ij} p_j. \end{cases} \tag{2.7}$$

It remains to check that the right-hand side coincides with the Hamiltonian vector field of H_g. In local canonical coordinates the Hamiltonian H_g is given by

$$H_g(q,p) = \frac{1}{2} g^{ij}(q) p_i p_j.$$

Its differential reads

$$dH_g = \frac{1}{2} g^{ij}_{,k} p_i p_j dq^k + g^{ij} p_j dp_i.$$

Hence the Hamiltonian vector field of H_g with respect to the symplectic form $\omega = dp_i \wedge dq^i$ equals

$$X_{H_g} = -\frac{1}{2} g^{ij}_{,k} p_i p_j \frac{\partial}{\partial p_k} + g^{ij} p_j \frac{\partial}{\partial q^i}.$$

In particular, the flow of X_{H_g} is given by the solution of (2.7). This finishes the proof of the theorem. □

2.3.2 Stereographic projection and the geodesic flow of the round metric

We discuss a simple, but important example of a geodesic flow, namely the geodesic flow of the round sphere, i.e., the sphere with the standard Riemannian metric with constant sectional curvature 1. We consider the standard embedding of the n-sphere in $\mathbb{R}^{n+1}$ as the subset

$$S^n = \{q \in \mathbb{R}^{n+1} \mid \|q\| = 1\},$$

and use stereographic projection from the north pole $N = (1, 0, \dots, 0)$. This projection is given by the map

$$\pi : S^n - N \longrightarrow \mathbb{R}^n, \quad (q_0; \vec{q}) \longmapsto \frac{\vec{q}}{1 - q_0}.$$

We briefly compute how the round metric looks like in stereographic coordinates. For this, we need the Jacobian of the inverse to π, which parametrizes the whole sphere except for the north pole N. This inverse is given by

$$\pi^{-1} : \mathbb{R}^n \longrightarrow S^n - N, \quad \vec{x} \mapsto \left(\frac{|\vec{x}|^2 - 1}{|\vec{x}|^2 + 1}; \frac{2\vec{x}}{|\vec{x}|^2 + 1} \right).$$

Its Jacobian is an $(n+1) \times n$ matrix, whose entries are given by

$$d(\pi^{-1})_{0,j} = \frac{4x_j}{(|\vec{x}|^2 + 1)^2}, \quad d(\pi^{-1})_{i,j} = \frac{2}{|\vec{x}|^2 + 1}\delta_{i,j} - \frac{4x_i x_j}{(|\vec{x}|^2 + 1)^2}.$$

We now compute the metric,

$$\begin{aligned} g_{ij} &= \left\langle d(\pi^{-1}) \cdot \frac{\partial}{\partial x_i}, d(\pi^{-1}) \cdot \frac{\partial}{\partial x_j} \right\rangle = \left(d(\pi^{-1})^t d(\pi^{-1}) \right)_{ij} \\ &= \sum_{k=0}^{n} d(\pi^{-1})_{k,i} d(\pi^{-1})_{k,j} \\ &= 4 \left(4 \frac{x_i x_j}{(|\vec{x}|^2 + 1)^4} + \frac{1}{(|\vec{x}|^2 + 1)^2} \sum_{k=1}^{n} \left(\delta_{k,i} - \frac{2x_k x_i}{|\vec{x}|^2 + 1} \right) \left(\delta_{k,j} - \frac{2x_k x_j}{|\vec{x}|^2 + 1} \right) \right) \\ &= \frac{4}{(|\vec{x}|^2 + 1)^2} \left(\delta_{i,j} + \frac{4x_i x_j}{(|\vec{x}|^2 + 1)^4} - \frac{4x_i x_j}{(|\vec{x}|^2 + 1)^2} + \frac{4x_i x_j \vec{x}^2}{(|\vec{x}|^2 + 1)^4} \right). \end{aligned}$$

After cleaning up this expression we obtain the matrix

$$g_{ij} = \frac{4\delta_{i,j}}{(|\vec{x}|^2 + 1)^2}. \tag{2.8}$$

We now find the Hamiltonian for the geodesic flow on the round sphere in stereographic coordinates, namely

$$H = \frac{1}{2} \sum_{i,j} g^{ij} p_i p_j = \frac{1}{8} (\|q\|^2 + 1)^2 \|p\|^2. \tag{2.9}$$

2.3.3 Mechanical Hamiltonians

Assume that (N, g) is a Riemannian manifold and $V \in C^\infty(N, \mathbb{R})$ is a smooth function on the configuration space referred to as the *potential*. We define the Hamiltonian

$$H_{g,V} : T^*N \to \mathbb{R}, \quad (q, p) \mapsto \frac{1}{2}|p|_g^2 + V(q),$$

i.e., in the language of physics the sum of kinetic and potential energy. Let us make the Hamiltonian vector field of such a Hamiltonian explicit in the special case where N is an open subset of $\mathbb{R}^n$ endowed with its standard scalar product.

In the following we will omit the inner product from the notation. With respect to the splitting $T^*\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$, the Hamiltonian vector field is given by

$$X_{H_V}(q,p) = \begin{pmatrix} p \\ -\nabla V(q) \end{pmatrix},$$

where ∇V is the gradient of V, which in terms of physics can be thought of as the negative of a *force*. Given (q,p) in $T^*N \subset T^*\mathbb{R}^n$, and if $q_p \colon \mathbb{R} \to N$ is a solution of the second-order ODE

$$\partial_t^2 q_p(t) = -\nabla V(q_p(t))$$

meeting the initial conditions

$$q_p(0) = q, \quad \partial_t q_p(0) = p,$$

then the Hamiltonian flow of H_V applied to (q,p) is given by

$$\phi^t_{H_V}(q,p) = (q_p(t), \partial_t q_p(t)).$$

We discuss three basic examples of mechanical Hamiltonians. The first example is the *harmonic oscillator*. Its Hamiltonian is given by

$$H\colon T^*\mathbb{R} \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}(p^2 + q^2).$$

Its flow is given by

$$\phi^t_H(q,p) = (q\cos t + p\sin t, -q\sin t + p\cos t).$$

Note that the flow of the harmonic oscillator is periodic with period 2π. We will often use the more practical complex notation, and write instead

$$\phi^t_H(p+iq) = e^{it}(p+iq).$$

Our second example is the case of *two uncoupled harmonic oscillators.* In this example, we fix two positive real numbers a_1 and a_2, and the Hamiltonian is given by

$$H\colon T^*\mathbb{R}^2 \to \mathbb{R}, \quad (q,p) \mapsto \frac{a_1}{2}(p_1^2 + q_1^2) + \frac{a_2}{2}(p_2^2 + q_2^2).$$

Note that in this example the Hamiltonian flow is just the product flow of two harmonic oscillators, possibly with different periods, with respect to the splitting $T^*\mathbb{R}^2 = T^*\mathbb{R} \times T^*\mathbb{R}$. We first specialize to the case $a_1 = a_2 = 1$. The flow is then again periodic of period 2π. Moreover, note that if $c > 0$, then the level set or energy hypersurface of two uncoupled harmonic oscillators $H^{-1}(c)$ is a three-dimensional sphere of radius $\sqrt{2c}$.

Slightly more interesting dynamics arise if a_1 and a_2 are distinct. In this case, the level sets $H^{-1}(c)$ become ellipsoids for $c > 0$. Furthermore, the flow is now given by

$$\phi_H^t(p_1 + iq_1, p_2 + iq_2) = (e^{ia_1 t}(p_1 + iq_1), e^{ia_2 t}(p_2 + iq_2)\,).$$

If a_1 is a rational multiple of a_2, let us assume in the following that a_1 and a_2 are positive integers, then the flow is periodic, but the periods differ. Orbits with $p_1 + iq_1 = 0$ have period $2\pi/a_2$, orbits with $p_2 + iq_2 = 0$ have period $2\pi/a_1$, and other orbits have period $2\pi/\gcd(a_1, a_2)$. When a_1 and a_2 are rationally independent, we obtain the *irrational ellipsoid*, which has, up to a time-shift, only two periodic orbits, namely the orbits $\gamma_1(t) = (e^{ia_1 t}, 0)$ and $\gamma_2(t) = (0, e^{ia_2 t})$.

Our third example is the *Kepler problem*. In this case the Hamiltonian is given by

$$H\colon T^*(\mathbb{R}^n \setminus \{0\}) \to \mathbb{R}, \quad (q, p) \mapsto \frac{1}{2}|p|^2 - \frac{1}{|q|}.$$

From the physical point of view the most relevant cases are the case $n = 2$, the *planar Kepler problem* and the case $n = 3$. Using Moser regularization we will see that for negative energy the Hamiltonian flow of the Kepler problem in any dimension n can be embedded up to reparametrization in the geodesic flow of the round n-dimensional sphere. In the planar case the double cover of the geodesic flow on the round two-dimensional sphere can be interpreted as the Hamiltonian flow of two uncoupled harmonic oscillators via the so-called Levi-Civita regularization, which we will discuss in Section 4.2.

2.3.4 Magnetic Hamiltonians

Mechanical Hamiltonians model physical systems where the force just depends on the position. There are, however, important forces which depend on the velocity as well. Examples are the Lorentz force in the presence of a magnetic field or the Coriolis force. To model such more general systems we twist the kinetic energy with a one-form. The set-up is as follows. Assume that (N, g) is a Riemannian manifold, $V \in C^\infty(N, \mathbb{R})$ is a potential, and in addition $A \in \Omega^1(N)$ is a one-form on N. We consider the Hamiltonian

$$H_{g,V,A}\colon T^*N \to \mathbb{R}, \quad (q, p) \mapsto \frac{1}{2}|p - A_q|^2 + V(q).$$

Because of its importance in the study of electromagnetism, such Hamiltonians are referred to as magnetic Hamiltonians.

2.3.5 Physical symmetries

The last class of Hamiltonians is less directly associated to Hamiltonian systems arising in physical situations. However, their Hamiltonian flows generate a family of physical transformations which are important to study symmetries of Hamiltonian systems. Assume that N is a manifold and $X \in \Gamma(TN)$ is a vector field on N. We associate to the vector field X a Hamiltonian

$$H_X \colon T^*N \to \mathbb{R}$$

as follows. Denote by $\pi \colon T^*N \to N$ the footpoint projection. We interpret a point $e \in T^*N$ as a vector $e \in T^*_{\pi(e)}N$. We can then pair this vector with the vector $X(\pi(e)) \in T_{\pi(e)}N$. Hence we set

$$H_X(e) := e\big(X(\pi(e))\big).$$

Assume that the flow $\phi^t_X \colon N \to N$ exists for every $t \in \mathbb{R}$. Then the Hamiltonian flow of H_X exists as well and is given by

$$\phi^t_{H_X} = d_*\phi^t_X, \tag{2.10}$$

where the symplectomorphism $d_*\phi$ for a diffeomorphism ϕ was defined in (2.1).

A prominent example of such a Hamiltonian is *angular momentum*. Namely, consider on $\mathbb{R}^2$ the vector field

$$X = q_1 \frac{\partial}{\partial q_2} - q_2 \frac{\partial}{\partial q_1} \in \Gamma(T\mathbb{R}^2).$$

Define angular momentum as the Hamiltonian

$$L \colon T^*\mathbb{R}^2 \to \mathbb{R}$$

given for $(q,p) \in T^*\mathbb{R}^2$ by

$$L(q,p) := H_X(q,p) = p_2 q_1 - p_1 q_2.$$

Note that the vector field X generates the counterclockwise rotation, i.e.,

$$\phi^t_X = R_t$$

where

$$R_t \colon \mathbb{R}^2 \to \mathbb{R}^2$$

is the counterclockwise rotation

$$R_t\big(q_1, q_2\big) = \big((\cos t)q_1 - (\sin t)q_2, (\sin t)q_1 + (\cos t)q_2\big).$$

Therefore in view of (2.10) we obtain for the Hamiltonian flow of angular momentum

$$\phi^t_L = d_* R_t \colon T^*\mathbb{R}^2 \to T^*\mathbb{R}^2. \tag{2.11}$$

2.3.6 Normal forms

The rich structure of the symplectomorphism group allows us to prove various neighborhood theorems. We will state one of these neighborhood theorems here, namely Darboux' theorem.

Theorem 2.3.2 (Darboux). *Assume that (M, ω) is a $2n$-dimensional symplectic manifold and $x \in M$. Then there exists an open neighborhood U of x and a diffeomorphism*

$$\Phi \colon U \to V \subset \mathbb{R}^{2n}$$

such that $\Phi(x) = 0 \in V$ and

$$\Phi^* \left(\sum_{i=1}^{n} dp_i \wedge dq_i \right) = \omega.$$

A proof using the famous Moser trick can be found in [189, Theorem 1.18].

2.4 Hamiltonian structures

A Hamiltonian manifold is the odd-dimensional analog of a symplectic manifold.

Definition 2.4.1. A *Hamiltonian manifold* is a pair (Σ, ω), where Σ is an odd-dimensional manifold, and $\omega \in \Omega^2(\Sigma)$ is a closed two-form with the property that $\ker \omega$ defines a one-dimensional distribution in $T\Sigma$. The two-form ω is called a *Hamiltonian structure* on Σ.

Here is how Hamiltonian manifolds arise in nature. Suppose that (M, ω) is a symplectic manifold and $H \in C^\infty(M, \mathbb{R})$ is a Hamiltonian with the property that 0 is a regular value of H. Consider the energy hypersurface

$$\Sigma = H^{-1}(0) \subset M.$$

It follows that the pair $(\Sigma, \omega|_\Sigma)$ is a Hamiltonian manifold. Moreover, it follows that in this example the one-dimensional distribution $\ker \omega$ is given by

$$\ker \omega = \langle X_H|_\Sigma \rangle, \tag{2.12}$$

i.e., the line bundle spanned by the restriction of the Hamiltonian vector field to Σ. Indeed, note that if $x \in \Sigma$ and $\xi \in T_x\Sigma$, then it holds that

$$\omega(X_H, \xi) = -dH(\xi) = 0,$$

where the last equality follows since ξ is tangent to a level set of H. In particular, the leaves of the distribution $\ker \omega|_\Sigma$ correspond to the trajectories of the flow ϕ_H^t restricted to Σ. However, by studying the leaves of $\ker \omega|_\Sigma$ instead of the flow of $\phi_H^t|_\Sigma$ we lose the information about their parametrization. One can say that people studying Hamiltonian manifolds instead of Hamiltonian systems are quite relaxed, since they do not care about time. This is often an advantage. Indeed, in order to regularize a Hamiltonian system one has, in general, to reparametrize the flow of the Hamiltonian. Therefore it is convenient to have a notion which remains invariant under these transformations.

Given a Hamiltonian manifold (Σ, ω) a *Hamiltonian vector field* is a non-vanishing section of the line bundle $\ker \omega$. Again if $\Sigma = H^{-1}(0)$ arises as the regular level set of a Hamiltonian function on a symplectic manifold, then the restriction of X_H to Σ is a Hamiltonian vector field. Note that if X is a Hamiltonian vector field on (Σ, ω), it follows from Cartan's formula that the Lie derivative $\mathcal{L}_X \omega = 0$, so that ω is preserved under the flow of X.

2.5 Contact forms

Definition 2.5.1. Assume that (Σ, ω) is a Hamiltonian manifold of dimension $2n-1$. A *contact form* for (Σ, ω) is a one-form $\lambda \in \Omega^1(\Sigma)$ which meets the following two assumptions

(i) $d\lambda = \omega$,

(ii) $\lambda \wedge \omega^{n-1}$ is a volume form on Σ.

Not every Hamiltonian manifold (Σ, ω) does admit a contact form. An obvious necessary condition for the existence of a contact form is that $[\omega] = 0 \in H^2_{dR}(\Sigma)$. We will call the pair (Σ, λ) a *contact manifold*. The defining property for λ is then the assumption

$$\lambda \wedge (d\lambda)^{n-1} > 0, \tag{2.13}$$

i.e., $\lambda \wedge (d\lambda)^{n-1}$ is a volume form. Each contact manifold becomes a Hamiltonian manifold by setting $\omega = d\lambda$.

Remark 2.5.2. In the literature, a contact manifold is more commonly defined as a manifold together with a maximally non-integrable field of hyperplanes, the contact structure. Any 1-form that locally defines this field of hyperplanes satisfies Equation (2.13). A more precise name for the notion we just defined would be a cooriented contact manifold: namely the hyperplane field $\xi = \ker \lambda$ is cooriented by the 1-form λ.

We will explain the notion of contact structure in somewhat more detail afterwards, but our main interest is in dynamics, and for this a contact form is the fundamental object.

Given a contact manifold (Σ, λ), the Reeb vector field $R \in \Gamma(T\Sigma)$ is implicitly defined by the conditions

$$\iota_R d\lambda = 0, \quad \lambda(R) = 1.$$

It follows that the Reeb vector field is a non-vanishing section in the line bundle $\ker d\lambda = \ker\omega \subset T\Sigma$. In particular, it is a Hamiltonian vector field of the Hamiltonian manifold (Σ, ω) and we have

$$\ker\omega = \langle R\rangle. \tag{2.14}$$

If Σ arises as the level set $\Sigma = H^{-1}(0)$ of a Hamiltonian H on a symplectic manifold, it follows from (2.12) and (2.14) that the Reeb vector field and the restriction of the Hamiltonian vector field $X_H|_\Sigma$ are parallel. In particular, their flows coincide up to reparametrization.

On the contact manifold (Σ, λ) we can further define the hyperplane field

$$\xi := \ker\lambda \subset T\Sigma.$$

This leads to a splitting

$$T\Sigma = \xi \oplus \langle R\rangle.$$

Note that the restriction of $d\lambda$ to ξ makes ξ a symplectic vector bundle of rank $2n-2$ over Σ.

The hyperplane distribution ξ is referred to as the *contact structure*. While the contact structure ξ is determined by the contact form λ, the converse does not hold. Indeed, if $f > 0$ is any positive smooth function on Σ, we obtain a new contact form

$$\lambda_f := f\lambda \in \Omega^1(\Sigma).$$

That λ_f is indeed a contact form can be checked by computing

$$\lambda_f \wedge (d\lambda_f)^{n-1} = f^n(\lambda \wedge (d\lambda)^{n-1}) > 0.$$

The two contact forms λ and λ_f give rise to the same contact structure

$$\xi = \ker\lambda = \ker\lambda_f.$$

On the other hand, the Reeb vector fields of λ and λ_f are in general not parallel to each other. Therefore the Reeb dynamics of the contact manifold (Σ, λ_f) might be quite different from the Reeb dynamics of the contact manifold (Σ, λ). This explains also that we cannot recover the Hamiltonian structure $\omega = d\lambda$ from the contact structure $\xi = \ker\lambda$. The study of contact manifolds is nowadays an interesting topic in its own right, see for example the book by Geiges [97]. In contrast to contact topology, our major concern is the Hamiltonian structure $\omega = d\lambda$, since this structure determines the dynamics up to reparametrization. From our point of view the contact form is more an auxiliary structure, which enables us to get information on the dynamics of the Hamiltonian manifold (Σ, ω). Indeed, the contact form turns out to be an indispensable tool in order to apply holomorphic curve techniques. Furthermore, we will see that with the help of contact forms one can rule out blue sky catastrophes.

We will not need many tools from contact topology, but the following statement, known as *Darboux's Theorem*, is fundamental, and will be used several times.

Theorem 2.5.3 (Darboux)**.** *Assume that* (N, λ) *is a three-dimensional contact manifold and* $p \in N$*. Then there exists an open neighborhood* U *of* p *and a diffeomorphism*

$$\Phi \colon U \to V \subset \mathbb{R}^3$$

such that $\Phi(p) = 0 \in V$ *and*

$$\Phi^*(dq_1 + q_2 dq_3) = \lambda.$$

The neighborhood from this theorem is often called Darboux chart or Darboux ball if V is $\mathbb{R}^3$. A proof using the Moser trick can be found in [97, Theorem 2.5.1]. This theorem is often rephrased by saying that a contact manifold has no local invariants or that every contact 3-manifold locally looks like Figure 2.1.

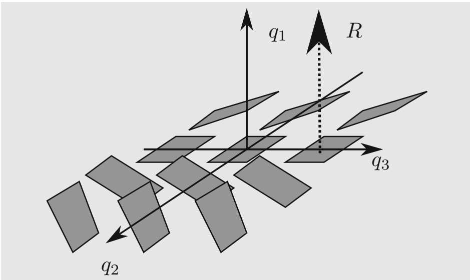


Figure 2.1: The standard contact structure on $\mathbb{R}^3$, with contact form $\lambda = dq_1 + q_2 dq_3$, and its Reeb field $R = \frac{\partial}{\partial q_1}$.

It is maybe helpful to point out that this theorem can also be seen as a special flow box theorem for the Reeb vector field. Indeed, in a Darboux chart the Reeb vector field is given by $\frac{\partial}{\partial q_1}$.

2.6 Liouville domains and contact type hypersurfaces

Contact manifolds can sometimes be obtained as some nice hypersurfaces in a symplectic manifold (M, ω). To be more precise, we consider a hypersurface $\Sigma \subset M$, and assume that there is a vector field X defined in a neighborhood of Σ satisfying

$$\mathcal{L}_X \omega = \omega. \tag{2.15}$$

In other words, the symplectic form ω is expanding along flow lines of X. So if ϕ_X^t denotes the flow of X, then

$$(\phi_X^t)^* \omega = e^t \omega. \tag{2.16}$$

We will call a vector field X satisfying (2.15) a *Liouville vector field*. If X is a Liouville vector field for ω, then we obtain a 1-form λ by $\lambda = \iota_X \omega$. This 1-form is called the Liouville form. We claim that

Proposition 2.6.1. *Suppose that X is a Liouville vector field defined on a neighborhood of a hypersurface $\Sigma \subset M$. Assume that X is transverse to Σ, so $T_x\Sigma \oplus \langle X_x \rangle = T_xM$ for all $x \in \Sigma$. Then $(\Sigma, (\iota_X\omega)|_\Sigma)$ is a contact manifold with contact form $(\iota_X\omega)|_\Sigma$.*

Proof. To see this, abbreviate $\lambda = \iota_X\omega$. Given $x \in \Sigma$, choose a basis $\{v_1, \ldots, v_{2n-1}\}$ of $T_x\Sigma$. We compute

$$\begin{aligned} \lambda \wedge (d\lambda)^{n-1}(v_1, \ldots, v_{2n-1}) &= \iota_X\omega \wedge \omega^{n-1}(v_1, \ldots, v_{2n-1}) \\ &= \frac{1}{n}\omega^n(X_x, v_1, \ldots, v_{2n-1}). \end{aligned}$$

Because X is transverse to Σ, it follows that $\{X_x, v_1, \ldots, v_{2n-1}\}$ is a basis of T_xM. Because ω is non-degenerate, it follows that

$$\omega^n(X_x, v_1, \ldots, v_{2n-1}) \neq 0$$

and we see that $\lambda|_\Sigma$ is indeed a contact form on Σ. □

A hypersurface Σ satisfying the assumptions of the proposition is called a *contact-type hypersurface*. We will now define a class of symplectic manifolds that come equipped with a contact type hypersurface. To explain this notion we assume that (M, λ) is an exact symplectic manifold, i.e., $\omega = d\lambda$ is a symplectic structure on M. Recall that an exact symplectic manifold cannot be closed, since for a closed symplectic manifold $[\omega] \neq 0 \in H^2_{dR}(M)$. We assume that M is compact, so it must have a non-empty boundary.

Definition 2.6.2. By a *Liouville domain* we mean a compact, exact symplectic manifold (M, λ) with the property that the Liouville vector field X, which is defined by $\iota_X d\lambda = \lambda$, is transverse to the boundary and outward pointing.

By Proposition 2.6.1 it follows that the boundary of a Liouville domain (M, λ) is contact with contact form $\lambda|_{\partial M}$, so we have

Lemma 2.6.3. *Assume that (M, λ) is a Liouville domain. Then $(\partial M, \lambda|_{\partial M})$ is a contact manifold.*

Given a Liouville domain (M, λ) with Liouville vector field X, we claim that

$$\mathcal{L}_X\lambda = \lambda \tag{2.17}$$

where $\mathcal{L}_X$ denotes the Lie derivative in direction of X. To see this we first observe that

$$\iota_X\lambda = \omega(X, X) = 0$$

by the antisymmetry of the form ω. Hence we compute using Cartan's formula

$$\mathcal{L}_X\lambda = d\iota_X\lambda + \iota_X d\lambda = \iota_X\omega = \lambda.$$

This proves (2.17). We give two examples of Liouville domains which play an important role in the following.

Example 2.6.4. The cotangent bundle $M = T^*N$ together with the Liouville one-form is an example of an exact symplectic manifold. In local canonical coordinates the Liouville one-form is given by $\lambda = \sum p_i dq_i$ and therefore the associated Liouville vector field reads

$$X = \sum_{i=1}^{n} p_i \frac{\partial}{\partial p_i}.$$

Now suppose that $\Sigma \subset T^*N$ is *fiberwise star-shaped*, i.e., for every $x \in N$ it holds that $\Sigma \cap T_x^*N$ bounds a star-shaped domain D_x in the vector space T_x^*N. Then $X \pitchfork \Sigma$ and $D = \bigcup_{x\in N} D_x$ is a Liouville domain with $\partial D = \Sigma$.

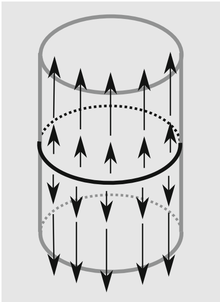


Figure 2.2: The cylinder, (T^*S^1, pdq), admits the Liouville field $p\frac{\partial}{\partial p}$.

Remark 2.6.5. If $H\colon T^*N \to \mathbb{R}$ is a mechanical Hamiltonian, i.e., $H(q,p) = \frac{1}{2}|p|_g^2 + V(q)$ for a Riemannian metric g on N and a smooth potential $V\colon N \to \mathbb{R}$, and $c > \max V$, then the energy hypersurface $\Sigma = H^{-1}(c)$ is fiberwise star-shaped. Indeed,

$$dH(X)(q,p) = |p|_g^2$$

and because $c > \max V$ it follows that p does not vanish on Σ so that we get

$$dH(X)|_\Sigma > 0.$$

Example 2.6.6. On the complex vector space $\mathbb{C}^n$ with coordinates $(z_1, \ldots, z_n) = (x_1 + iy_1, \ldots, x_n + iy_n)$ consider the one-form

$$\lambda = \frac{1}{2}\sum_{i=1}^{n}(x_i dy_i - y_i dx_i).$$

The pair $(\mathbb{C}^n, \lambda)$ is a (linear) symplectic manifold with symplectic form

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i$$

and Liouville vector field

$$X = \frac{1}{2}\sum_{i=1}^{n}\left(x_i\frac{\partial}{\partial x_i} + y_i\frac{\partial}{\partial y_i}\right).$$

If $\Sigma \subset \mathbb{C}^n$ is *star-shaped*, i.e., Σ bounds a star-shaped domain D, then $X \pitchfork \Sigma$ and D is a Liouville domain with $\partial D = \Sigma$.

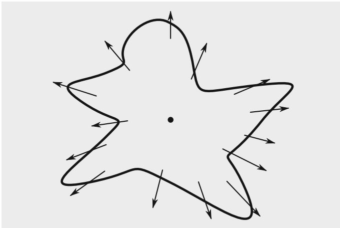


Figure 2.3: A star-shaped domain in $\mathbb{C}$.

2.7 Real Liouville domains and real contact manifolds

A *real Liouville domain* is a triple (M, λ, ρ) where (M, λ) is a Liouville domain and $\rho \in \mathrm{Diff}(M)$ is an exact anti-symplectic involution, i.e.,

$$\rho^2 = \mathrm{id}|_M, \quad \rho^*\lambda = -\lambda.$$

Because the exterior derivative commutes with pullback, we immediately obtain that

$$\rho^*\omega = -\omega$$

for the symplectic form $\omega = d\lambda$, i.e., ρ is an anti-symplectic involution. It follows that the Liouville vector field X defined by $\iota_X\omega = \lambda$ is invariant under ρ, meaning

$$\rho^* X = X.$$

The fixed point set $\mathrm{Fix}(\rho)$ of an anti-symplectic involution is a (maybe empty) Lagrangian submanifold. To see this, pick $x \in \mathrm{Fix}(\rho)$. The differential

$$d\rho(x)\colon T_xM \to T_xM$$

is then a linear involution. Therefore we have the vector space decomposition

$$T_xM = \ker(d\rho(x) - \mathrm{id}) \oplus \ker(d\rho(x) + \mathrm{id})$$

into the eigenspaces of $d\rho(x)$ to the eigenvalues ± 1. By choosing a ρ-invariant Riemannian metric on M, we see that, locally around x, the fixed point set $\mathrm{Fix}(\rho)$

can be parametrized by the restriction of the exponential map to $\ker(d\rho(x) - \mathrm{id})$, the eigenspace to the eigenvalue 1. This proves that $\mathrm{Fix}(\rho)$ is a submanifold of M and moreover,

$$T_x\mathrm{Fix}(\rho) = \ker(d\rho(x) - \mathrm{id}).$$

Because ρ is anti-symplectic, both eigenspaces are isotropic subspaces of the symplectic vector space T_xM, i.e., ω vanishes on both of them. Hence by dimensional reasons they have to be Lagrangian, i.e., isotropic subspaces of the maximal possible dimension, namely half the dimension of M. This proves that $\mathrm{Fix}(\rho)$ is Lagrangian. If ρ is an exact anti-symplectic involution then in addition the restriction of λ to $\mathrm{Fix}(\rho)$ vanishes as well, so that $\mathrm{Fix}(\rho)$ becomes an *exact Lagrangian submanifold.*

Example 2.7.1. On $\mathbb{C}^n$ complex conjugation is an involution under which the Liouville form $\lambda = \frac{1}{2}\sum_{i=1}^n (x_i dy_i - y_i dx_i)$ is anti-invariant. Therefore if $\Sigma \subset \mathbb{C}^n$ is a star-shaped hypersurface invariant under complex conjugation, then Σ bounds a real Liouville domain. Note that the fixed point set of complex conjugation is the Lagrangian $\mathbb{R}^n \subset \mathbb{C}^n$. This explains the terminology real Liouville domain.

Example 2.7.2. Let N be a closed manifold and $I \in \mathrm{Diff}(N)$ a smooth involution, i.e., $I^2 = \mathrm{id}|_N$. By (2.2) the physical transformation $d_*I\colon T^*N \to T^*N$ is an exact symplectic involution on T^*N, namely

$$(d_*I)^*\lambda = \lambda$$

for the Liouville one-form λ. Consider furthermore the involution $\rho_1\colon T^*N \to T^*N$ which on every fiber restricts to minus the identity so that in canonical coordinates we have

$$\rho_1(q,p) = (q,-p).$$

The involution ρ_1 is an exact anti-symplectic involution which commutes with the exact symplectic involution

$$\rho_1 \circ d_*I = d_*I \circ \rho_1 =: \rho_I$$

so that ρ_I is an exact anti-symplectic involution on T^*N. Note that

$$\mathrm{Fix}(\rho_I) = \nu^*\mathrm{Fix}(I)$$

is the conormal bundle of the fixed point set of I. Indeed, a point $(q,p) \in T^*N$ lies in $\mathrm{Fix}(\rho_I)$ precisely when

$$q = I(q), \quad dI(q)p = -p.$$

This means that (q,p) lies in the conormal bundle of $\mathrm{Fix}(I)$. If $\Sigma \subset T^*N$ is a fiberwise star-shaped hypersurface invariant under ρ_I, then Σ bounds a real Liouville domain. In particular, if we take I the identity on N, then the anti-symplectic involution is just $\rho_{\mathrm{id}} = \rho_1$ and therefore the requirement is that Σ is fiberwise symmetric star-shaped. Note that a mechanical Hamiltonian is invariant under the involution ρ_1, so that according to Remark 2.6.5 the energy hypersurface of a mechanical Hamiltonian for energies higher than the maximum of the potential bounds a real Liouville domain.

Now assume that (M, λ, ρ) is a real Liouville domain with boundary $\Sigma = \partial M$. Denote by abuse of notation the restrictions of λ and ρ to Σ by the same letter. Then the triple (Σ, λ, ρ) is a *real contact manifold*, namely a contact manifold (Σ, λ) together with an anti-contact involution $\rho \in \mathrm{Diff}(\Sigma)$, namely an involution satisfying $\rho^*\lambda = -\lambda$. The Reeb vector field $R \in \Gamma(T\Sigma)$ is then anti-invariant under ρ, i.e.,

$$\rho^* R = -R.$$

The fixed point set of an anti-contact involution is a (maybe empty) *Legendrian submanifold* of Σ, namely a submanifold whose tangent bundle is a Lagrangian subbundle of the symplectic vector bundle $\xi = \ker(\lambda)$. Moreover, the flow $\phi_R^t \colon \Sigma \to \Sigma$ for $t \in \mathbb{R}$ of the Reeb vector field satisfies the following relation with ρ,

$$\phi_R^t = \rho \circ \phi_R^{-t} \circ \rho. \tag{2.18}$$

In order to see this, we compute

$$\phi_R^t = \phi_{-R}^{-t} = \phi_{\rho^* R}^{-t} = \rho^{-1} \circ \phi_R^{-t} \circ \rho = \rho \circ \phi_R^{-t} \circ \rho$$

where for the last equality we have used that ρ is an involution.

Chapter 3

Symmetries

3.1 Poisson brackets and Noether's theorem

In order to simplify differential equations, it is important to identify *preserved quantities*, also called integrals. More formally, if X is a vector field on a manifold M, then we call L an *integral* of X if $X(L) = 0$.

The notion of Poisson bracket will be helpful. For a symplectic manifold (M, ω) we define the *Poisson bracket* of smooth functions F and G by

$$\{F, G\} := \omega(X_F, X_G) = -dF(X_G) = dG(X_F) = -X_G(F) = X_F(G). \tag{3.1}$$

We see directly from the definition that the Poisson bracket describes the time-evolution of a function. Indeed, suppose that $\gamma(t)$ is a flow line of X_F. Then

$$\frac{dG \circ \gamma(t)}{dt} = X_F(G) = \{F, G\}.$$

From this identity, energy preservation, see Theorem 2.2.2, follows because $\{H, H\} = 0$ (the Poisson bracket is alternating). Before we turn our attention to conserved quantities, we first need to establish some properties of the Poisson bracket.

Lemma 3.1.1. *Given smooth functions F, G on a symplectic manifold (M, ω), we have the following relation between the Lie bracket and Poisson bracket,*

$$[X_F, X_G] = X_{\{F,G\}}.$$

Proof. We first rewrite the Lie bracket a bit:

$$[X_F, X_G] = \mathcal{L}_{X_F} X_G = \frac{d}{dt}|_{t=0} {\phi^t_{X_F}}^* X_G = \frac{d}{dt}|_{t=0} X_{G \circ \phi^t_{X_F}}.$$

For the last equality, note that Theorem 2.2.3 tells us that the flow of X_F preserves the symplectic form, so that its flow pulls back the Hamiltonian vector field of G

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_3

to the Hamiltonian vector field of the pull back of the Hamiltonian G. Now use this identity and the definition:

$$\begin{aligned} i_{[X_F,X_G]}\omega &= \frac{d}{dt}|_{t=0}\omega(X_{G\circ\phi^t_{X_F}},\cdot) \\ &= \frac{d}{dt}|_{t=0}\left(-d(G\circ\phi^t_{X_F}\right) \\ &= -d\left(\frac{d}{dt}|_{t=0}G\circ\phi^t_{X_F}\right) = -d\left(X_F(G)\right) = -d\{F,G\}. \end{aligned}$$ □

Apply Theorem 2.3.2 to find a Darboux chart $(U, \omega = dp\wedge dq)$ for (M,ω). Then we can write the Poisson bracket as

$$\{F,G\} = \sum_i \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q^i} - \frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i}.$$

In a moment, we shall see that the Poisson bracket endows the space of smooth functions on M with a Lie algebra structure. We briefly recall the definition.

Definition 3.1.2. A *Lie algebra* consists of a vector space $\mathfrak{g}$ together with a binary operation $[\cdot,\cdot]$ that is bilinear, alternating, so $[v,v]=0$ for all $v\in\mathfrak{g}$ and satisfies the Jacobi identity, i.e., for all $X,Y,Z\in\mathfrak{g}$, we have

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0.$$

Proposition 3.1.3. *The pair $(C^\infty(M),\{\cdot,\cdot\})$ is a Lie algebra.*

Proof. We check the required properties. First of all, note that for $a\in\mathbb{R}$ and $F,G\in C^\infty(M)$, we have $X_{aF+G} = aX_F + X_G$, since the Hamiltonian vector field is the solution to a linear equation. Hence, $\{aF+G,H\} = \omega(X_{aF+G},X_H) = a\omega(X_F,X_H) + \omega(X_G,X_H) = a\{F,H\} + \{G,H\}$. The same argument works for the second factor, so $\{\cdot,\cdot\}$ is bilinear. Also, $\{F,F\} = \omega(X_F,X_F) = 0$, so $\{\cdot,\cdot\}$ is alternating. Alternatively, we can also use Lemma 3.1.1. Finally, we check that the Jacobi identity holds by computing the individual terms:

$$\begin{aligned} \{F,\{G,H\}\} &= X_F(\{G,H\}) = X_F(X_G(H)), \\ \{G,\{H,F\}\} &= -X_G(\{F,H\}) = -X_G(X_F(H)), \\ \{H,\{F,G\}\} &= -X_{\{F,G\}}(H) = -[X_F,X_G](H). \end{aligned}$$

We have used Lemma 3.1.1 in the last step. Summing these terms shows that the Jacobi identity holds. □

Lemma 3.1.4. *The function G is an integral of X_F if and only if $\{F,G\}=0$.*

Proof. The function G is an integral if and only if $X_F(G)=0$. This holds if and only if $0 = -dF(X_G) = \omega(X_F,X_G) = \{F,G\}$. □

Remark 3.1.5. By Lemma 3.1.1 $\{F, G\} = 0$ implies that $[X_F, X_G] = 0$. On the other hand, for G to be an integral of X_F, it is not enough to just have $[X_F, X_G] = 0$. Indeed, consider $(\mathbb{R}^2, \omega_0 = dp \wedge dq)$ with the Hamiltonians $F = p$ and $G = q$. Then $X_F = \partial_q$ and $X_G = -\partial_p$, so $[X_F, X_G] = 0$. However, G is linearly increasing under the flow of X_F, so G is not an integral of X_F. On the other hand the next lemma shows that if M is closed such a phenomenon cannot happen.

Lemma 3.1.6. *Assume that (M, ω) is a closed symplectic manifold and $F, G \in C^\infty(M, \mathbb{R})$ are two smooth functions such that $[X_F, X_G] = 0$. Then $\{F, G\} = 0$.*

Proof. Because the commutator of the two Hamiltonian vector fields vanishes, it follows from Lemma 3.1.1 that

$$X_{\{F,G\}} = [X_F, X_G] = 0.$$

We assume without loss of generality that M is connected (otherwise we treat each connected component of M separately). Therefore we conclude that

$$\{F, G\} = c$$

where $c \in \mathbb{R}$ is a constant. Pick $x \in M$. To study the behavior of G along the flow of X_F through x we differentiate

$$\frac{d}{dt} G\big(\phi_F^t(x)\big) = dG\big(\phi_F^t(x)\big) X_F\big(\phi_F^t(x)\big) = \{F, G\}\big(\phi_F^t(x)\big) = c.$$

We conclude that

$$G\big(\phi_F^t(x)\big) = G(x) + ct.$$

Because M is compact by assumption, the function G is bounded and therefore

$$c = 0.$$

This proves that F and G Poisson commute. □

We are now in position to prove Noether's theorem.

Theorem 3.1.7 (Noether)**.** *Assume that (M, ω) is a closed symplectic manifold and $F, G \in C^\infty(M, \mathbb{R})$. Then the following are equivalent.*

(i) *G is an integral for the flow of F, i.e., $G(\phi_F^t(x))$ is independent of t for every $x \in M$.*
(ii) *The flow of G is a symmetry for F, i.e., $F(\phi_G^t(x))$ is independent of t for every $x \in M$.*
(iii) *F and G Poisson commute, i.e., $\{F, G\} = 0$.*
(iv) *The flows of X_F and X_G commute, i.e., $[X_F, X_G] = 0$.*

Proof. Lemma 3.1.4 tells us that assertion (i) is equivalent to assertion (iii). Because the Poisson bracket is antisymmetric, the vanishing of $\{F, G\}$ is equivalent to the vanishing of $\{G, F\}$. Therefore assertion (iii) is equivalent to assertion (ii). Lemma 3.1.1 and Lemma 3.1.6 imply that assertion (iii) is equivalent to assertion (iv). □

3.2 Hamiltonian group actions and moment maps

Consider a smooth action of a Lie group G on a symplectic manifold (M, ω). Take a vector $\xi \in \mathfrak{g} = \mathrm{Lie}(G)$, the Lie algebra of G. Then we get a path $\exp(t\xi)$ in G and act on M with this path. Take the derivative with respect to t to get a vector field on M, namely

$$X_\xi(x) := \left.\frac{d}{dt}\right|_{t=0} (\exp(t\xi) \cdot x). \tag{3.2}$$

If we assume that the action preserves the symplectic form ω, then X_ξ is a symplectic vector field, i.e., a vector field having the property that its Lie derivative of the symplectic form ω vanishes, and therefore equivalently its flow preserves the symplectic form.

Remark 3.2.1. Note that X_ξ does not need to be a Hamiltonian vector field. For instance the vector field ∂_θ on $(T^2, d\theta \wedge d\phi)$ is symplectic, yet not Hamiltonian.

We will now assume that X_ξ is Hamiltonian, so for each X_ξ there is a function H_ξ satisfying $\iota_{X_\xi}\omega = -dH_\xi$. Then we get a map

$$\begin{aligned} \psi : \mathfrak{g} &\longrightarrow C^\infty(M) \\ \xi &\longmapsto H_\xi. \end{aligned}$$

Note that the Hamiltonian is determined by its Hamiltonian vector field just up to addition of constants on each connected component of the manifold. Therefore ψ is only determined by the action of G up to addition of a function from the Lie algebra to the reals on any connected component of M. We make the following definition.

Definition 3.2.2. We say the action is *Hamiltonian* if the map ψ can be chosen to be a Lie algebra homomorphism.

There are obstructions for actions to be Hamiltonian, but this will not be relevant here. Instead, we point out a couple of examples.

Example 3.2.3. We act with an $SO(n)$-matrix A on $T^*\mathbb{R}^n$ by the formula $A\cdot(q,p) = (Aq, Ap)$, or in other words, the standard action on each component. This action preserves the symplectic form.

The Lie algebra $\mathfrak{so}(n)$ can be identified with skew-symmetric matrices together with the usual commutator as Lie bracket.

Given a skew-symmetric $n \times n$-matrix B, we define the Hamiltonian

$$H_B(q,p) := p^t Bq.$$

The Hamiltonian equations are

$$\dot{p} = -\partial_q H_B = (-p^t B)^t = -B^t p = Bp, \quad \dot{q} = \partial_p H_B = Bq,$$

which we can solve by exponentiating, so $(q(t), p(t)\,) = (e^{Bt}q, e^{Bt}p)$. This is precisely the $SO(n)$-action on $T^*\mathbb{R}^n$.

Theorem 3.2.4 (Hamiltonian version of Noether's theorem)**.** *Suppose that G is a Lie group acting Hamiltonianly on a symplectic manifold (M, ω). If $H : M \to \mathbb{R}$ is a Hamiltonian that is invariant under G, then each $\xi \in \mathfrak{g}$ gives an integral H_ξ of X_H, or equivalently $\{H, H_\xi\} = 0$.*

Proof. Take $\xi \in \mathfrak{g}$. We get a vector field X_ξ on M by formula (3.2), which is the Hamiltonian vector field of the function H_ξ by the assumption of Hamiltonian action. The Hamiltonian H is assumed to be invariant, so by the formula for the Poisson bracket we have

$$0 = X_{H_\xi}(H) = -X_H(H_\xi),$$

so H_ξ is an integral of X_H. □

Given a Hamiltonian group action we define a smooth map

$$\mu \colon M \to \mathfrak{g}^*$$

where $\mathfrak{g}^*$ is the dual of the Lie algebra $\mathfrak{g}$ by

$$\langle \mu(x), \xi \rangle = H_\xi(x), \quad x \in M,\ \xi \in \mathfrak{g}.$$

Assume now that G is connected. We claim that the map μ is equivariant with respect to the action of G on M and the coadjoint action of G on $\mathfrak{g}^*$, i.e.,

$$\langle \mu(gx), \xi \rangle = \langle \mu(x), \mathrm{Ad}(g)\xi \rangle, \quad x \in M,\ \xi \in \mathfrak{g},\ g \in M. \tag{3.3}$$

Remark 3.2.5. As a reminder, we recall that the adjoint action of a Lie group G on its Lie algebra $\mathfrak{g}$, denoted by $\mathrm{Ad}(g)$, is the derivative of conjugation. In other words,

$$\mathrm{Ad}(g)\xi = \frac{d}{dt} g^{-1} \exp(t\xi) g|_{t=0}.$$

The adjoint representation of $\mathfrak{g}$ in $\mathfrak{gl}(\mathfrak{g})$, denoted by $\mathrm{ad}(\eta)$, is in turn the derivative of this adjoint action, meaning

$$\mathrm{ad}(\eta)\xi = \frac{d}{dt} \mathrm{Ad}(\exp(t\eta))\xi|_{t=0}.$$

This can be rewritten as $\mathrm{ad}(\eta)\xi = [\eta, \xi]$.

Because G is assumed to be connected, this is equivalent to the infinitesimal version of (3.3). Namely, the requirement

$$\big\langle d\mu(x) X_\eta(x), \xi \big\rangle = \big\langle \mu(x), \mathrm{ad}(\eta)\xi \big\rangle, \quad x \in M,\ \xi, \eta \in \mathfrak{g}. \tag{3.4}$$

For the left-hand side we compute

$$\big\langle d\mu(x) X_\eta(x), \xi \big\rangle\ = dH_\xi(x) X_\eta(x) = dH_\xi(x) X_{H_\eta}(x) = \{H_\eta, H_\xi\}(x),$$

while for the right-hand side we obtain

$$\big\langle \mu(x), \mathrm{ad}(\eta)\xi \big\rangle = H_{\mathrm{ad}(\eta)\xi}(x) = H_{\{\eta,\xi\}}(x)$$

and therefore (3.4) follows from the assumption that $\psi\colon \mathfrak{g} \to C^\infty(M)$ is a Lie algebra homomorphism. This proves (3.3).

The map $\mu\colon M \to \mathfrak{g}^*$ is known as *moment map* or *momentum map*. By the discussion above it is characterized by the following two properties

(i) μ is equivariant with respect to the action of G on M and the coadjoint action of G on $\mathfrak{g}^*$.

(ii) For every $\xi \in \mathfrak{g}$ it holds that $X_{H_\xi} = X_{\langle \mu,\xi\rangle}$.

The notion of moment map goes back to the work of Souriau [224, 225]. We refer to [170] for more information on the history of moment maps. More information on the rich topic of moment maps can be found for example in [100, 172, 195].

Example 3.2.6. The following example generalizes Example 3.2.3. Assume that a Lie group G acts on a manifold N. We obtain a symplectic action of G on T^*N, namely if $g \in G$ and $e \in T^*N$, then

$$g_* e = d_* g(e).$$

A moment map for this action can be constructed as follows. For $\xi \in \mathfrak{g} = \mathrm{Lie}(G)$ denote by X_ξ^N the infinitesimal generator of the action of G on N, i.e., $X_\xi^N(q) := \frac{d}{dt}\big|_{t=0}(\exp(t\xi)q)$ for $q \in N$. Then the moment map $\mu\colon T^*N \to \mathfrak{g}$ is defined as

$$\big\langle \mu(e), \xi \big\rangle = \big\langle e, X_\xi^N(\pi(e)) \big\rangle, \quad e \in T^*N,\ \xi \in \mathfrak{g}$$

where $\pi\colon T^*N \to N$ is the footpoint projection.

3.3 Angular momentum, the spatial Kepler problem, and the Runge–Lenz vector

3.3.1 Central force: conservation of angular momentum

Suppose we are given a Hamiltonian dynamical system H on $T^*\mathbb{R}^3$. Define the *angular momentum* by

$$L := q \times p,$$

where $(q,p) \in \mathbb{R}^3 \times \mathbb{R}^3 = T^*\mathbb{R}^3$. Consider the Hamiltonian

$$H = \frac{1}{2}|p|^2 + V(|q|) \tag{3.5}$$

on $(\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}^n$, where $V : \mathbb{R} \to \mathbb{R}$ is (smooth) function, possibly with some singularities. Such a function V is called the potential for a *central force*, because it only depends on the distance.

We will assume that $n = 3$, although this can be generalized.

Lemma 3.3.1. *The angular momentum is preserved under the flow of X_H. In other words, the components of the angular momentum $L = (L_1, L_2, L_3)$ satisfy $\{H, L_i\} = 0$.*

Proof. By Example 3.2.3, the standard $SO(n)$ action acts Hamiltonianly on $T^*\mathbb{R}^n$. The Hamiltonian for a central force is $SO(n)$-invariant, so Noether's theorem, Theorem 3.2.4, implies the claim. □

Remark 3.3.2. The physical interpretation of preservation of angular momentum is that flow lines of the Hamiltonian vector field X_H lie in the plane with normal vector L.

3.3.2 The Kepler problem and its integrals

We shall consider the Hamiltonian

$$H = \frac{1}{2}|p|^2 - \frac{1}{|q|} \tag{3.6}$$

on $(\mathbb{R}^n \backslash \{0\}) \times \mathbb{R}^n$ with coordinates (q, p) and standard symplectic form $\omega = dp \wedge dq$. The latter is shorthand notation for

$$dp \wedge dq = \sum_i dp_i \wedge dq^i.$$

The physically relevant case is $n = 2, 3$, and we shall first consider the case $n = 3$. The equations of motion are

$$\begin{aligned} \dot{p} &= -\frac{q}{|q|^3} \\ \dot{q} &= p. \end{aligned}$$

In other words, the force equals $\ddot{q} = -\frac{q}{|q|^3}$, so its strength drops off with the distance squared.

The strategy is to find as many integrals as possible, and in fact the Kepler problem turns out to be *completely integrable*. We will define this notion later, but roughly speaking, we can say that this means that there are so many conserved quantities that we can convert the ODE's into algebraic equations.

Lemma 3.3.3. *The angular momentum L is an integral of the Kepler problem.*

The Kepler problem has an obvious $SO(3)$-symmetry, or put differently the force is central, so Lemma 3.3.1 applies.

Remark 3.3.4. In higher dimensions this Hamiltonian has an $SO(n)$-symmetry, but we will not consider this more general (and unphysical) situation.

3.3.3 The Runge–Lenz vector: another integral of the Kepler problem

The following integral depends on the specific form of the Kepler Hamiltonian (3.6). Define the *Laplace–Runge–Lenz vector* (also called *Runge–Lenz vector*)[1] by

$$A := p \times L - \frac{q}{|q|}.$$

Lemma 3.3.5. *The Runge–Lenz vector A is preserved under the flow of X_H. In other words, the components of $A = (A_1, A_2, A_3)$ satisfy $\{H, A_i\} = 0$.*

Remark 3.3.6. Unlike the preservation of angular momentum, this integral is not obvious from a symmetry of the phase space, which makes this definition appear to fall from the sky. However, we will show in the next chapter that the spatial Kepler problem can be transformed into the geodesic flow of the round metric on S^3, which shows that the spatial Kepler problem actually comes with an $SO(4)$-symmetry. Similarly, the planar Kepler problem has an $SO(3)$-symmetry. We also point out that the Runge–Lenz vector can be a priori geometrically motivated.

We will prove that A is an integral by a short, elementary computation.

Proof. We compute the time-derivative of A,

$$\begin{aligned}
\dot{A} &= \dot{p} \times L + p \times \dot{L} - \frac{\dot{q}}{|q|} + \frac{q}{|q|^2}\frac{q \cdot \dot{q}}{|q|} \\
&= -\frac{q}{|q|^3} \times (q \times p) - \frac{p}{|q|} + \frac{q}{|q|^3}(q \cdot p) \\
&= \frac{1}{|q|^3}\left(-q \times (q \times p) - (q \cdot q)p + (q \cdot p)q\right) = 0.
\end{aligned}$$

In the second step we have used the Hamilton equations and that $\dot{L} = 0$, and in the last step we used the vector product identity

$$(u \times v) \times w = (u \cdot w)v - (v \cdot w)u. \qquad \square$$

Lemma 3.3.7. *The Runge–Lenz vector satisfies the identity*

$$|A|^2 = 1 + 2H \cdot |L|^2. \tag{3.7}$$

Proof. The following computation makes use of the fact that p and L are orthogonal and the identity $q \cdot p \times L = \det(q, p, L) = q \times p \cdot L$. We find

$$\begin{aligned}
|A|^2 &= |p \times L|^2 - \frac{2}{|q|} q \cdot p \times L + \frac{|q|^2}{|q|^2} = 1 + |p|^2|L|^2 - \frac{2}{|q|}|L|^2 \\
&= 1 + 2\left(\frac{1}{2}|p|^2 - \frac{1}{|q|}\right)|L|^2. \qquad \square
\end{aligned}$$

[1] This integral was discovered by Jakob Hermann (1678–1733). We refer to [102, 103] for the history of this vector.

Solving the Kepler problem

Define the plane $P_L = \{v \in \mathbb{R}^3 \mid \langle L, v\rangle = 0\}$.

Lemma 3.3.8. *The vector A lies in the plane P_L.*

Recall that $\langle q, L\rangle = 0$ and observe that

$$\langle A, L\rangle = \langle p \times L, L\rangle - \langle \frac{q}{|q|}, L\rangle = 0 + 0.$$

To describe the motion of the particle more explicitly, we apply a coordinate

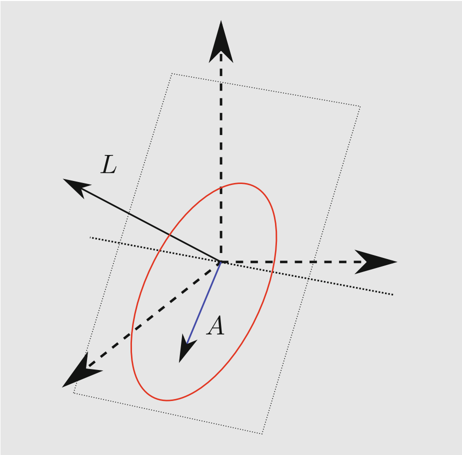


Figure 3.1: A sketch of a Kepler orbit in the plane P_L.

change, namely a rotation to move the L-vector to the z-axis. Then $L = (0, 0, \ell)$ for some $\ell > 0$, and hence we can write

$$A = (|A| \cos g, |A| \sin g, 0).$$

Definition 3.3.9. The angle g is called the *argument of the perigee* (*perihelion*)[2].

We now determine the radius as function of the angle ϕ. Using the above formula for A and the identity $\langle p \times L, q\rangle = \det(p, L, q)$, we find

$$|q| + \langle A, q\rangle = \langle \frac{q}{|q|}, q\rangle + \langle A, q\rangle = \langle p \times L, q\rangle = \det(p, L, q) = \langle q \times p, L\rangle = |L|^2.$$

As before, we write q in polar coordinates

$$q = (r \cos \phi, r \sin \phi, 0)$$

[2]Perigee means close to the Earth. Perihelion means close to the Sun. If the heavy mass describes the Earth, one uses perigee, if it is the Sun, one uses the word perihelion.

and by plugging this into $|q| + \langle A, q\rangle = |L|^2$, we find

$$r = \frac{|L|^2}{1 + |A|\cos(\phi - g)}. \tag{3.8}$$

It is common to call the quantity $f := \phi - g$ the *true anomaly*, and $|A|$ is called the *eccentricity*. The geometric picture is indicated in Figure 3.1. In non-collision orbits, where $L \neq 0$, we see from Equation (3.8) that the perigee, the closest approach to Earth, by which we mean the point on the orbit closest to the origin, is attained at $\phi = g$, the argument of the perigee.

With the above computations, we can deduce the following classification result for solutions.

Theorem 3.3.10. *Solutions to the Kepler problem are conic sections, i.e., curves that are the intersection of a plane and a cone.*

Proof. We intersect the plane given by $z = |L|^2 - |A| \cdot x$ with the cone given by $z = \sqrt{x^2 + y^2}$. This gives the equation

$$\sqrt{x^2 + y^2} = |L|^2 - |A| \cdot x,$$

which is equivalent to the above equation for a solution of the Kepler problem,

$$|q| + \langle A, q\rangle = |L|^2$$

if we substitute $\sqrt{x^2 + y^2} = |q|$ and $\langle A, q\rangle = |A|r\cos\phi = |A| \cdot x$. □

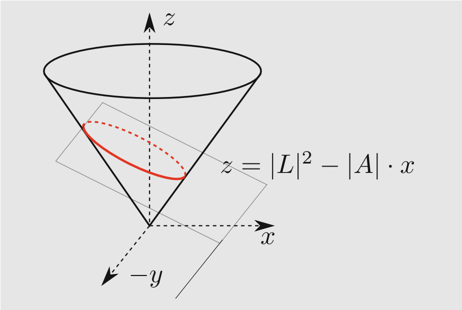


Figure 3.2: Solutions to the Kepler problem are conic sections.

3.4 Completely integrable systems

A completely integrable Hamiltonian system on a $2n$-dimensional symplectic manifold is a Hamiltonian dynamical system with n Poisson-commuting, independent integrals. The following theorem tells us that this is enough to completely understand the flow. This is in contrast with a more general smooth dynamical system, where one would expect to need $2n-1$ integrals in order to confine the dynamics to a one-dimensional submanifold.

Theorem 3.4.1 (Arnold–Liouville: action-angle coordinates)**.** *Suppose that $\{H_i\}_{i=1}^n$ are n Poisson-commuting integrals on a symplectic manifold (M^{2n},ω).*

We suppose that the vector $\vec{h}=(h_1,\dots,h_n)$ is a regular value of the map $\vec{H}: x \mapsto (H_1(x),\dots,H_n(x))$. Assume in addition that $L := \vec{H}^{-1}(\vec{h})$ is connected and compact. Then the following hold.

- *The set L is a Lagrangian torus, known as* Liouville torus.
- *There is a neighborhood $\nu_M(L)$ that is diffeomorphic to $T^n \times D^n$ via the diffeomorphism*

$$\Phi : T^n \times D^n \longrightarrow \nu_M(L).$$

- *There are coordinates ϕ, S on $T^n \times D^n$ with the properties*
  1. *The symplectic form is standard:*
  $$\Phi^*\omega = \sum_j dS_j \wedge d\phi^j.$$
  2. *The coordinates S_j depend only on the integrals $\{H_k\}_{k=1}^n$.*
  3. *The flow $\phi_{H_j}^t$ is linear in these coordinates, i.e., $\Phi^{-1}\circ\phi_{H_j}^t\circ\Phi(\phi,S) = (\phi+tq(S),S)$.*

The coordinates constructed by this theorem are known as *action-angle coordinates*.

Below we give a proof which follows the proof in the book of Bolsinov and Fomenko closely, [40, Section 1.4]. Other references are Arnold's book, the book by Arnold, Kozlov, and Neishtadt, the book of Hofer and Zehnder and the book of Cushman and Bates, [18, 20, 131, 66]. We also mention the papers by Mineur [179, 180] where the first known proof of this result appeared.

Proof. We define the map

$$\begin{aligned}\Phi : \mathbb{R}^n \times M &\longrightarrow M \\ (\vec{t},x) &\longmapsto \phi_{X_{H_n}}^{t_n}\circ\cdots\circ\phi_{X_{H_1}}^{t_1}(x).\end{aligned}$$

We will consider the effect of this map on L, and write $\Phi|_L : \mathbb{R}^n \times L \to L$ for the restricted map. Note that $[X_{H_i},X_{H_k}]=0$ by the Noether theorem, so these flows commute, giving rise to an $\mathbb{R}^n$-action on L. Clearly, the map $\Phi|_L : \mathbb{R}^n \times L \to L$ is surjective, so we can write L as the disjoint union of orbits. Suppose that O_p is one such orbit of which L is comprised. We claim O_p is an open set in L. Indeed, if q lies in O_p, then the open neighborhood obtained from $(B_\epsilon(0), q)$ also lies in O_p, so q is an interior point.

Since L is connected by assumption, it follows that L consists of only one orbit, namely O_x. In other words, the $\mathbb{R}^n$-action is transitive.

We now use that L is compact. Since the $\mathbb{R}^n$-action on L is transitive, we find that L is diffeomorphic to

$$L \cong \mathbb{R}^n/\Lambda,$$

where Λ is the isotropy subgroup of a point, also called stabilizer subgroup. This group must be a discrete subgroup.

Lemma 3.4.2. *If Λ is a discrete subgroup of $\mathbb{R}^n$, then Λ is a lattice.*

By compactness of L it follows that the lattice Λ has full rank, namely $\operatorname{rk}\Lambda = n$, so L is indeed a torus. To see that L is a Lagrangian torus, we note that $\{X_{H_i}\}_i$ forms a basis of the tangent space. Furthermore, by the assumption that the integrals H_i Poisson-commute, we find

$$\omega(X_{H_i}, X_{H_j}) = \{H_i, H_j\} = 0,$$

so $\omega|_L = 0$, i.e., L is a Lagrangian torus.

We now produce the required coordinates. First we claim that the lattice vector $\lambda(h) \in \Lambda$ depends smoothly on the value of h. To see this, we apply the implicit function theorem. Indeed, in a chart near (x, h) we can solve the equation $\Phi(\lambda; x, h) - (x, h) = 0$ by the implicit function theorem giving $\lambda = \lambda(x, h)$. Furthermore, $\Phi(\cdot, h)$ preserves the value of h, since it is the composition of the flows of commuting Hamiltonian flows, so we find a smooth function $\lambda = \lambda(h)$ such that $\Phi(\lambda(h); x, h) = (x, h)$. In other words, the lattice depends smoothly on h. Fix now a smoothly varying basis of the lattice $\Lambda(h)$, which we denote by $\{\lambda_i(h)\}_{i=1}^n$.

Now pick a point $p(h)$ on each Lagrangian torus $L(h)$ depending smoothly on $h \in D^n_\epsilon$. If $x \in \nu_M(L)$, then we find $H(x) = (H_1(x), \ldots, H_n(x)) \in D^n_\epsilon$. To obtain the angular coordinates, we define $\psi(x) = (\psi^1(x), \ldots, \psi^n(x)) \in \mathbb{R}^n/(2\pi\mathbb{Z})^n$ by solving the equation

$$\phi_{H_n}^{\frac{\psi^n(x)}{2\pi}\lambda_n} \circ \cdots \circ \phi_{H_1}^{\frac{\psi^1(x)}{2\pi}\lambda_1}(p(H(x))) = x. \tag{3.9}$$

Since we defined the lattice $\Lambda(h)$ as the isotropy subgroup of the action, this is well defined, so the following smooth map

$$\nu_M(L) \longrightarrow \mathbb{R}^n/2\pi\mathbb{Z} \times \mathbb{R}^n, \quad x \longmapsto (\psi^1(x), \ldots, \psi^n(x); H_1(x), \ldots, H_n(x))$$

is a bijection whose differential has full rank, so they define coordinates. We will construct the desired action-angle coordinates from these coordinates.

First we observe that Equation (3.9) tells us that the tangent vectors $\frac{\partial}{\partial\psi^i}$ can be written as a linear combination of the Hamiltonian vector fields $\{X_{H_j}\}_{j=1}^n$ with coefficients depending only on the h-coordinates, so there are unique coefficients B such that

$$\frac{\partial}{\partial\psi^i} = \sum_j B^j{}_i(h)X_{H_j}.$$

This means that $\omega\left(\frac{\partial}{\partial H^k}, \frac{\partial}{\partial\psi^j}\right) = \sum_j B^j{}_i(h)dH_j\left(\frac{\partial}{\partial H^k}\right) = B^k{}_i$. Furthermore, $\omega|_L = 0$, and so the symplectic form in our coordinate system (H, ψ) has no terms of the form $d\psi^i \wedge d\psi^j$. We find

$$\omega = \sum_{i,j} A_{ij}dH_i \wedge dH_j + \sum_{k,\ell} B^k{}_\ell dH_k \wedge d\psi^\ell.$$

Since ω is closed and the Lie brackets $[\frac{\partial}{\partial H_i}, \frac{\partial}{\partial H_j}]$ and $[\frac{\partial}{\partial H_i}, \frac{\partial}{\partial\psi^k}]$ vanish, we find by the coordinate free version of the exterior differential that

$$\begin{aligned}
0 &= d\omega\left(\frac{\partial}{\partial H_i}, \frac{\partial}{\partial H_j}, \frac{\partial}{\partial\psi^k}\right)\\
&= \frac{\partial}{\partial H_i}\omega\left(\frac{\partial}{\partial H_j}, \frac{\partial}{\partial\psi^k}\right) - \frac{\partial}{\partial H_j}\omega\left(\frac{\partial}{\partial H_i}, \frac{\partial}{\partial\psi^k}\right) + \frac{\partial}{\partial\psi^k}\omega\left(\frac{\partial}{\partial H_i}, \frac{\partial}{\partial H_j}\right)\\
&= \frac{\partial}{\partial H_i}\left(\sum_\ell B^\ell{}_k(H)dH_\ell\left(\frac{\partial}{\partial H_j}\right)\right) - \frac{\partial}{\partial H_j}\left(\sum_\ell B^\ell{}_k(h)dH_\ell\left(\frac{\partial}{\partial H_i}\right)\right) + \frac{\partial}{\partial\psi^k}A_{ij}.
\end{aligned}$$

We see that

$$\frac{\partial}{\partial\psi_k}A_{ij} = \frac{\partial}{\partial h_j}\left(B^i{}_k(h)\right) - \frac{\partial}{\partial h_i}\left(B^j{}_k(h)\right). \tag{3.10}$$

Now we note that on one hand $\frac{\partial}{\partial\psi^k}A_{ij}$ is periodic in all ψ-directions, yet on the other hand, the right-hand side of (3.10) is independent of the ψ-coordinates. This implies that $\frac{\partial}{\partial\psi^k}A_{ij} = 0$, so the coefficients A_{ij} only depend on the H-coordinates.

If we put

$$\alpha := \sum_{i,j} A_{ij}dH^i \wedge dH^j \text{ and } \beta_\ell := \sum_k B^k{}_\ell dH^k,$$

then we conclude by the above that first of all $d\alpha = 0$, and secondly that $d\beta_\ell = 0$.

By restricting ourselves to a starshaped neighborhood $\nu_{D^n}(\vec{h})$ of the point $\vec{h}$, the Poincaré lemma gives us primitives of α and β_ℓ. We will write $\alpha = d\kappa$ and $\beta_\ell = ds_\ell$, and claim that the functions $\{s_\ell\}_\ell$ form coordinates on a possibly smaller neighborhood of $\vec{h}$. To see this, we apply the inverse function theorem after verifying that the Jacobian $\frac{\partial s_\ell}{\partial H_k}$ has full rank. By substituting the above primitives into our expression for the symplectic form, we find

$$\omega = d\kappa + \sum_\ell ds_\ell \wedge d\psi^\ell.$$

Since ω is non-degenerate, we have

$$\begin{aligned}0 \neq \omega^n &= n!(-1)^{\frac{n(n-1)}{2}} ds_1 \wedge \cdots \wedge ds_n \wedge d\psi^1 \wedge \cdots \wedge d\psi^n \\ &= n!(-1)^{\frac{n(n-1)}{2}} \det \left(\frac{\partial s_\ell}{\partial H_k} \right) dH_1 \wedge \cdots \wedge dH_n \wedge d\psi^1 \wedge \cdots \wedge d\psi^n,\end{aligned}$$

so we conclude that the Jacobian $\frac{\partial s_\ell}{\partial H_k}$ has indeed full rank. For the final step, first write κ in the new coordinates s_ℓ, i.e., $\kappa = \sum \kappa^i ds_i$, and put

$$\phi^i := \psi^i - \kappa^i.$$

Clearly, s_i and ϕ^j form coordinates on $\nu_M(L)$. These are action-angle coordinates as the following computation shows

$$\sum_i ds_i \wedge d\phi^i = \sum_i ds_i \wedge d\psi^i + \sum_i d\kappa^i \wedge ds_i = \sum_i \beta_i \wedge d\psi^i + d\kappa = \omega. \qquad \square$$

Remark 3.4.3. The name "action" comes from the following. On a cotangent bundle T^*N with completely integrable system $\{H_i\}_i$ consider the action of a loop γ_i with $\dot\gamma_i = \frac{\partial}{\partial \psi^i}$ in the sense of physics. This action is given by

$$\mathcal{A}(\gamma_i) = \int_{\gamma_i} pdq - Hdt.$$

This corresponds to the Rabinowitz functional which we describe in Equation (7.1) of Chapter 7. To keep things simple, assume that C_i is a disk capping the loop γ_i. After shifting the Hamiltonian by a constant, we can assume that $\int_{\gamma_i} Hdt = 0$, and we find by applying Stokes that

$$\mathcal{A}(\gamma_i) = \int_{\gamma_i} pdq - \int_{\gamma_i} s_j d\psi^j + \int_{\gamma_i} s_j d\psi^j = \int_{C_i} (\omega - \omega) + 2\pi s_i.$$

So the action along the loop γ_i is just the value of the action variable s_i after rescaling by 2π. Without the strong assumptions in this remark, the action is, in general, multi-valued.

Remark 3.4.4. The above theorem tells us that the flow of completely integrable systems is very simple in Arnold–Liouville coordinates. This makes completely integrable systems very special. They nevertheless play an important role, because

- we can understand their dynamics well.
- many important systems are close to integrable. For example, the restricted three-body problem can be written as

$$H_0 + \sum_{k=1}^{\infty} \mu^k H_k,$$

where H_0 is the rotating Kepler problem, a completely integrable system.

3.5 The planar Kepler problem

In Section 3.3.2 we discussed the spatial Kepler problem, so the planar problem follows as a special case. Later on we are mostly interested in the planar restricted three-body problem. Therefore it will be helpful to develop some explicit formulas for the planar problem.

The Hamiltonian for the planar Kepler problem is given by

$$E\colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|^2 - \frac{1}{|q|}. \tag{3.11}$$

We prefer to abbreviate the Hamiltonian for the Kepler problem by E, as an abbreviation for energy and not by H as usual. This is because we also have to study the Kepler problem in rotating coordinates, the so-called rotating Kepler problem, and we prefer to save the letter H to denote the Hamiltonian in rotating coordinates.

In the spatial Kepler problem angular momentum is a three-dimensional vector. In the planar case the first two components of this vector vanish and only the third component survives. If we talk about angular momentum for the planar Kepler problem we just mean this third component, hence

$$L\colon T^*\mathbb{R}^2 \to \mathbb{R}, \quad (q,p) \mapsto q_1 p_2 - q_2 p_1.$$

The Kepler problem is rotationally invariant. Because rotation is generated by angular momentum we obtain by Noether's theorem

$$\{E, L\} = 0, \tag{3.12}$$

as we discussed already in Lemma 3.3.3, a formula which the reader is invited to check directly. Because the phase space of the planar Kepler problem is four-dimensional and the integrals E and L are independent on a set of full measure, one says that the planar Kepler problem is a completely integrable system, which slightly abuses our restricted definition of this notion.

It is not hard to see in this example the invariant tori predicted in the Arnold–Liouville Theorem 3.4.1. Let us discuss this for negative energy. In this case orbits of the Kepler problem are either ellipses or collision orbits. An ellipse is topologically a circle, and if we rotate it we get a torus. An exception to this is the case where the ellipse is a circle. In this case by rotating it we just get the circle back. But this is no contradiction to the Arnold–Liouville theorem. Namely along the circle the Hamiltonian vector fields of E and L are parallel, so that a circle does not lie on a regular value of the map $(E,L)\colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}^2$. For collision orbits the level set is non-compact so that we cannot apply the Arnold–Liouville theorem directly, either. However, we will soon regularize collisions in Chapter 4 to get rid of this annoying non-compactness.

Apart from the physical symmetry obtained by rotation which is generated by the Hamiltonian vector field of angular momentum, the Kepler problem admits also some "hidden symmetries". These hidden symmetries do not arise from flows on the configuration space $\mathbb{R}^2 \setminus \{0\}$ but from flows which only live on phase space $T^*(\mathbb{R}^2 \setminus \{0\})$. We discussed the Runge–Lenz vector in Section 3.3.3. In the planar case the third component of this vector vanishes so that just the first two are of interest. They are the smooth functions

$$A_1, A_2 \colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}$$

given by

$$\begin{cases} A_1(q,p) = p_2(p_2q_1 - p_1q_2) - \frac{q_1}{|q|} = p_2L(q,p) - \frac{q_1}{|q|}, \\ A_2(q,p) = -p_1(p_2q_1 - p_1q_2) - \frac{q_2}{|q|} = -p_1L(q,p) - \frac{q_2}{|q|}. \end{cases}$$

By Lemma 3.3.5 the Poisson bracket of E with A_1 and A_2 vanishes, i.e.,

$$\{E, A_1\} = \{E, A_2\} = 0.$$

For the planar Kepler problem the two-dimensional vector

$$A = (A_1, A_2) \colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R} \tag{3.13}$$

is referred to as the *Runge–Lenz vector*. By Lemma 3.3.7 we have the equality

$$|A|^2 = 1 + 2EL^2, \tag{3.14}$$

from which the following inequality for energy and angular momentum in the Kepler problem follows

$$2EL^2 + 1 \geq 0. \tag{3.15}$$

As we have seen in Section 3.3.2 the length of the Runge–Lenz vector A corresponds to the eccentricity of the conic section. Therefore equality holds if and only if the trajectory lies on a circle.

We finally work out the Hamiltonian of the planar Kepler problem in polar coordinates and deduce Kepler's second law. In polar coordinates for the q_1q_2-plane, there is a nice expression of L. We write $(q_1, q_2) = (r\cos\phi, r\sin\phi)$. The momentum coordinates (p_x, p_y) transform with the inverse of the Jacobian, so if we denote the cotangent coordinates dual to (r, ϕ) by (p_r, p_ϕ), then we find

$$(p_x, p_y) = \left(\cos\phi \cdot p_r - \frac{\sin\phi}{r} p_\phi, \sin\phi \cdot p_r + \frac{\cos\phi}{r} p_\phi\right).$$

The coordinate change for the q-coordinates is

$$q_x = q_r \cos(q_\phi), \quad q_y = q_r \sin(q_\phi).$$

We are looking for a symplectic transformation, so, as mentioned, we just need the inverse and transpose of the Jacobian of this coordinate transformation for the

p-part. It is, however, convenient to compute by using the fact that the canonical 1-form, $\lambda = p_x dq_x + p_y dq_y$ is preserved. This gives the equation $p_x dq_x + p_y dq_y = p_r dq_r + p_\phi dq_\phi$, so we find $p_x = p_r \cos q_\phi - \frac{p_\phi}{q_r} \sin q_\phi$ and $p_y = p_r \sin q_\phi + \frac{p_\phi}{q_r} \cos q_\phi$. The Hamiltonian in polar coordinates is hence

$$E = \frac{1}{2}\left(p_r^2 + \frac{p_\phi^2}{q_r^2}\right) - \frac{1}{q_r}.$$

Remark 3.5.1. It is important to observe that p_ϕ is the angular momentum. The Hamiltonian equations clearly show that the angular momentum is preserved as we expect for a central force.

If one plugs this into $L = q_1 p_2 - q_2 p_1$ and uses the Hamilton equations $\dot{q}_i = p_i$, then one finds Kepler's second law.

Lemma 3.5.2 (Kepler's second law)**.** *We have*

$$\frac{1}{2} r^2 \dot{\phi} = \frac{1}{2} L = \frac{d\,\mathrm{Area}}{dt}.$$

where Area *is the area swept out by an ellipse.*

To get the claim about the area, just use that

$$Area = \int_{\phi=\phi_1}^{\phi_2} \int_{r=0}^{r=r_1} r dr d\phi = \int_{\phi=\phi_1}^{\phi_2} \frac{1}{2} r^2 d\phi.$$

We assume that $L \neq 0$. The case $L = 0$ can be worked out separately: it involves collision orbits.

Chapter 4

Regularization of Two-Body Collisions

4.1 Moser regularization

As we have seen in the discussion of the planar Kepler problem, the Kepler problem admits an obvious rotational symmetry, namely rotations of the plane, forming the group $SO(2)$, but also admits hidden symmetries, which are generated by the Runge–Lenz vector (3.13). These hidden symmetries played an important role in the early development of quantum mechanics, in particular in Pauli's and Fock's discussion about the spectrum of the hydrogen atom [87, 202]. In [186] Moser explained how the Kepler flow can be embedded into the geodesic flow of the round sphere. This explains the hidden symmetries because the round metric is invariant under the group $SO(3)$. In the case of the planar Kepler problem we obtain the geodesic flow on the two-dimensional sphere and the symmetry group becomes $SO(3)$. In particular, the symmetry group is three-dimensional, in accordance with the fact that we have three integrals, namely angular momentum as well as the two components of the Runge–Lenz vector. We refer to the works of Hulthén [137] and Bargmann [24] for a discussion of these symmetries in terms of quantum mechanics. How the Runge–Lenz vector is related to the moment map of the Hamiltonian action of $SO(3)$ on the cotangent bundle of the two-dimensional sphere is for example explained in [108, 150, 227].

We first explain Moser regularization of the Kepler problem at the energy value $-\frac{1}{2}$. Let $E\colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}$ be the Kepler Hamiltonian (3.11), i.e., the map $(q,p) \mapsto \frac{1}{2}|p|^2 - \frac{1}{|q|}$. Define

$$K(p,q) = \frac{1}{2}\bigg(|q|\Big(E(-q,p)+\frac{1}{2}\Big)+1\bigg)^2 = \frac{1}{2}\Big(\frac{1}{2}(|p|^2+1)|q|\Big)^2.$$

Recall from Section 2.2.2 the switch map which is symplectic and interchanges the roles of position and momentum. If we do these acrobatics in our mind, the Hamiltonian K is just the kinetic energy of the "momentum" q with respect to the round metric on S^2 in the chart obtained by stereographic projection, see Equation (2.9)

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_4

and the preceding computation for more details. Hence by Theorem 2.3.1 the flow of the Hamiltonian vector field of K is given by the geodesic flow of the round two-sphere in the chart obtained by stereographic projection. Note that

$$dK|_{E^{-1}(-\frac{1}{2})}(p,q) = |q| dE|_{E^{-1}(-\frac{1}{2})}(-q,p).$$

In particular, because the switch map $(p,q) \mapsto (-q,p)$ is a symplectomorphism, the Hamiltonian vector fields restricted to the energy hypersurface $E^{-1}(-\frac{1}{2}) = K^{-1}(\frac{1}{2})$ are related by

$$X_K|_{E^{-1}(-\frac{1}{2})}(p,q) = |q| X_E|_{E^{-1}(-\frac{1}{2})}(-q,p).$$

That means the flow of X_K is just a reparametrization of the flow of X_E. In particular, after reparametrization the Kepler flow at energy $-\frac{1}{2}$ can be interpreted as the geodesic flow of the round two-sphere in the chart obtained by stereographic projection.

The energy hypersurface $E^{-1}(-\frac{1}{2})$ is non-compact. This is due to *collisions*. However, the geodesic flow in the chart on the round two-sphere obtained by stereographic projection extends to the geodesic flow on the whole two-sphere. This procedure *regularizes* the Kepler problem. If we think of the two-sphere as $S^2 = \mathbb{R}^2 \cup \{\infty\}$, then the point at infinity corresponds to collisions. Indeed, at a collision point the original momentum p explodes which corresponds to the point at infinity of S^2 after interchanging the roles of momentum and position. It is also useful to note that the geodesic flow on the round two-sphere is completely periodic. The Kepler flow at negative energy is almost periodic. Trajectories are either ellipses which are periodic or collision orbits which are not periodic. Heuristically one can imagine that by regularizing the Kepler problem one builds in some kind of trampoline into the mass at the origin, such that when the body collides it just bounces back. The corresponding bounce orbit then becomes periodic in accordance with the fact that the geodesic flow on the round two-sphere is periodic.

To get a feeling what is going on at a more conceptual level it is useful to discuss regularization in terms of Hamiltonian manifolds. Recall that by going from the energy hypersurface of a Hamiltonian in a symplectic manifold to the underlying Hamiltonian manifold one loses the information on the parametrization of the trajectories. Because during the regularization procedure one has to reparametrize the trajectories, it is conceptually easier to forget the parametrization altogether and just discuss the whole procedure via Hamiltonian structures. Given a non-compact Hamiltonian manifold (Σ, ω) the question is if there exists a *closed* Hamiltonian manifold $(\overline{\Sigma}, \overline{\omega})$ and an embedding

$$\iota \colon \Sigma \to \overline{\Sigma} \quad \text{such that} \quad \iota^* \overline{\omega} = \omega.$$

Consider now more generally the Kepler problem at a negative energy value $c < 0$. In this case the Kepler flow is still up to reparametrization equivalent to the geodesic flow of the round two-sphere in stereographic projection. However the symplectic transformation $(p,q) \mapsto (-q,p)$ now has to be combined with the physical transformation $(q,p) \mapsto (\frac{q}{\sqrt{-2c}}, \sqrt{-2c}p)$. Note that

$$K(p,q) = \frac{1}{2}\bigg\{-\frac{|q|}{2c}\Big[E\Big(-\frac{q}{\sqrt{-2c}}, \sqrt{-2c}p\Big) - c\Big] + \frac{1}{\sqrt{-2c}}\bigg\}^2 = \frac{1}{2}\Big(\frac{1}{2}(|p|^2+1)|q|\Big)^2.$$

The two Hamiltonian vector fields on the energy hypersurface

$$\Sigma_c := E^{-1}(c) = K^{-1}(-\tfrac{1}{4c})$$

are related by

$$X_K|_{\Sigma_c}(p,q) = \frac{|q|}{(-2c)^{\frac{3}{2}}} X_E|_{\Sigma_c}\big(-\tfrac{q}{\sqrt{-2c}}, \sqrt{-2c}p\big)$$

respectively

$$X_E|_{\Sigma_c}(q,p) = -\frac{2c}{|q|} X_K|_{\Sigma_c}\big(\tfrac{p}{\sqrt{-2c}}, -\sqrt{-2c}q\big).$$

Suppose now that $\gamma \in C^\infty(S^1, \Sigma_c)$ is a Kepler ellipse, i.e., a solution of the ODE

$$\partial_t \gamma(t) = \tau X_E(\gamma(t)), \quad t \in S^1$$

where $\tau > 0$ is the minimal period of the ellipse. We next express the period τ in terms of the energy value c. In view of the discussion above we can interpret γ as a simple closed geodesic for the geodesic flow on the round two-sphere. In view of $\Sigma_c = K^{-1}(-\frac{1}{4c})$ the momentum of the geodesic has length $\sqrt{-2c}$. Therefore if $\lambda = -qdp$ is the Liouville one-form on the cotangent bundle of the sphere (note that momentum and position interchanged their role) the action of the simple closed geodesic is

$$\int_{S^1} \gamma^* \lambda = \frac{2\pi}{\sqrt{-2c}}. \tag{4.1}$$

Consider the one-form

$$\lambda' = \lambda - d(qp) = -2qdp - pdq.$$

Because λ and λ' only differ by an exact one-form, we have by Stokes

$$\int_{S^1} \gamma^* \lambda = \int_{S^1} \gamma^* \lambda'. \tag{4.2}$$

Using that the Hamiltonian vector field of the Kepler Hamiltonian is given by

$$X_E = p\frac{\partial}{\partial q} - \frac{q}{|q|^3}\frac{\partial}{\partial p}$$

we compute

$$\int_{S^1} \gamma^* \lambda' = \int_0^1 \lambda'(\tau X_E(\gamma))dt = \tau \int_0^1 \Big(-|p|^2 + \frac{2}{|q|} \Big) dt = -2c\tau. \tag{4.3}$$

Combining (4.1), (4.2), and (4.3) we obtain

$$\tau = -\frac{\pi}{c\sqrt{-2c}}.$$

We have proved the following version of Kepler's third law

Lemma 4.1.1 (Kepler's third law)**.** *The minimal period τ of a Kepler ellipse only depends on the energy and we have the relation*

$$\tau^2 = \frac{\pi^2}{-2E^3}.$$

Remark 4.1.2. It was proved by Belbruno [30] and Osipov [197, 198] that for energy 0 the Kepler flow after regularization becomes equivalent to the geodesic flow on Euclidean space and for positive energies it becomes equivalent to the geodesic flow on hyperbolic space, see also [31, 33, 98, 178].

4.2 The Levi-Civita regularization

We embed $\mathbb{C}$ into its cotangent bundle $T^*\mathbb{C}$ as the zero section. We get a smooth map

$$\mathscr{L} \colon \mathbb{C}^2 \setminus (\mathbb{C} \times \{0\}) \to T^*\mathbb{C} \setminus \mathbb{C}, \quad (u, v) \mapsto \Big(\frac{u}{\bar{v}}, 2v^2 \Big).$$

If we think of $\mathbb{C}$ as a chart of S^2 via stereographic projection at the north pole, the map $\mathscr{L}$ extends to a smooth map

$$\mathscr{L} \colon \mathbb{C}^2 \setminus \{0\} \to T^*S^2 \setminus S^2$$

which we denote by abuse of notation by the same letter. The map $\mathscr{L}$ is a covering map of degree 2. If we write (p, q) for coordinates of $T^*\mathbb{C} = \mathbb{C} \times \mathbb{C}$, where a bit unconventionally but justified by Moser regularization, we write p for the base coordinate and q for the fiber coordinate, the Liouville one-form on $T^*\mathbb{C}$ is given by

$$\lambda = q_1 dp_1 + q_2 dp_2 = \mathrm{Re}(q d\bar{p}).$$

If we pull back the Liouville one-form by the map $\mathscr{L}$ we obtain

$$\begin{aligned} \lambda_{\mathscr{L}}(u, v) &:= \mathscr{L}^* \lambda(u, v) = \mathrm{Re}\left(2v^2 d\left(\frac{\bar{u}}{v} \right) \right) = 2\mathrm{Re}\left(v^2 \left(\frac{d\bar{u}}{v} - \frac{\bar{u} dv}{v^2} \right) \right) \\ &= 2\mathrm{Re}\big(v d\bar{u} - \bar{u} dv \big) = 2\big(v_1 du_1 - u_1 dv_1 + v_2 du_2 - u_2 dv_2 \big). \end{aligned}$$

Its exterior derivative is the symplectic form

$$\omega_{\mathscr{L}} = 4\big(dv_1 \wedge du_1 + dv_2 \wedge du_2\big).$$

Note that $\lambda_{\mathscr{L}}$ does not agree with the standard Liouville one-form on $\mathbb{C}^2$ given by

$$\lambda_{\mathbb{C}^2} = \frac{1}{2}\big(u_1 du_2 - u_2 du_1 + v_1 dv_2 - v_2 dv_1\big)$$

and $\omega_{\mathscr{L}}$ differs from the standard symplectic form on $\mathbb{C}^2$

$$\omega_{\mathbb{C}^2} = d\lambda_{\mathbb{C}^2} = du_1 \wedge du_2 + dv_1 \wedge dv_2.$$

Indeed, the two subspaces $\mathbb{C} \times \{0\}$ and $\{0\} \times \mathbb{C}$ of $\mathbb{C}^2$ are Lagrangian with respect to $\omega_{\mathscr{L}}$ but symplectic with respect to $\omega_{\mathbb{C}^2}$. Nevertheless the Liouville vector field of $\lambda_{\mathscr{L}}$ implicitly defined by

$$\iota_{X_{\mathscr{L}}} \omega_{\mathscr{L}} = \lambda_{\mathscr{L}}$$

is given by

$$X_{\mathscr{L}} = \frac{1}{2}\left(u_1 \frac{\partial}{\partial u_1} + u_2 \frac{\partial}{\partial u_2} + v_1 \frac{\partial}{\partial v_1} + v_2 \frac{\partial}{\partial v_2}\right)$$

and agrees with the Liouville vector field of $\lambda_{\mathbb{C}^2}$. Recall that the standard Liouville vector field on T^*S^2 defined by $\iota_X d\lambda = \lambda$ for the standard Liouville one-form on T^*S^2 is given by

$$X = q\frac{\partial}{\partial q}$$

where q denotes the fiber variable. Because pull back commutes with exterior derivative we obtain

$$\mathscr{L}^* X = X_{\mathscr{L}}.$$

This implies the following lemma.

Lemma 4.2.1. *A closed hypersurface $\Sigma \subset T^*S^2$ is fiberwise star-shaped if and only if $\mathscr{L}^{-1}\Sigma \subset \mathbb{C}^2$ is star-shaped.*

Note that a fiberwise star-shaped hypersurface in T^*S^2 is diffeomorphic to the unit cotangent bundle S^*S^2 which itself is diffeomorphic to the three-dimensional projective space $\mathbb{R}P^3$. On the other hand a star-shaped hypersurface in $\mathbb{C}^2$ is diffeomorphic to the three-dimensional sphere S^3 which is a twofold cover of $\mathbb{R}P^3$.

In practice the Levi-Civita regularization of planar two-body collisions [160] is carried out by the variable substitution $(q,p) \mapsto (2v^2, \frac{u}{\bar{v}})$. As pointed out by Chenciner in [54] this substitution is already anticipated by Goursat [104], so it is much older than Moser regularization [186]. While Moser regularization works in every dimension, the Levi-Civita regularization depends on the existence of complex numbers which only exist in dimension two. Using quaternions instead of complex numbers an analog of the Levi-Civita regularization can be constructed in the spatial case [156].

We illustrate how the Levi-Civita regularization works for the Kepler problem at the energy value $-\frac{1}{2}$. We consider the Hamiltonian

$$H\colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}, \quad (q,p) \mapsto \frac{|p|^2}{2} - \frac{1}{|q|} + \frac{1}{2}.$$

After variable substitution we obtain

$$H(u,v) = \frac{|u|^2}{2|v|^2} - \frac{1}{2|v|^2} + \frac{1}{2}.$$

We introduce the Hamiltonian

$$K(u,v) := |v|^2 H(u,v) = \frac{1}{2}\big(|u|^2 + |v|^2 - 1\big).$$

The level set

$$\Sigma := H^{-1}(0) = K^{-1}(0)$$

equals the three-dimensional sphere. The Hamiltonian flow of K on Σ is just a reparametrization of the Hamiltonian flow of H on Σ and is periodic. Physically it can be interpreted as the flow of two uncoupled harmonic oscillators. Summarizing we have seen the following. If we apply Moser regularization to the Kepler problem of negative energy, we obtain the geodesic flow of S^2 and the regularized energy hypersurface becomes $\mathbb{R}P^3$. Its double cover is S^3 which we directly obtain by applying the Levi-Civita regularization to the Kepler problem. The double cover of the geodesic flow on S^2 can be interpreted as the Hamiltonian flow of two uncoupled harmonic oscillators.

4.3 Ligon–Schaaf regularization

Although Moser regularization regularizes each negative energy level separately and requires a reparametrization of time it is possible to regularize all negative energy levels at once without reparametrizing time. Such a regularization can be found in the paper by Györgyi [108] which itself was influenced by the work of Bacry, Ruegg, and Souriau [22] or the paper by Souriau [227]. The explicit description of such regularization map is due to Ligon and Schaaf [163], see also the paper by Cushman and Duistermaat [67]. How the regularization of Ligon–Schaaf is related to Moser regularization is explained in the paper by Heckman and de Laat in [112]. The regularization of Ligon–Schaaf embeds all negative energy levels of the Kepler problem into the cotangent bundle of the two-dimensional sphere.

A serious drawback of this regularization is, however, that the inverse is *not* smooth at collision points. In addition, there is no easy explicit description of the inverse of this embedding, which is of course related to the previous problem.

As a result, this regularization has not yet been used effectively to regularize the restricted three-body problem.

On the other hand, a nice feature of this embedding is that it interchanges the symmetries of the planar Kepler problem coming from angular momentum as well as the Runge–Lenz vector with the $SO(3)$-symmetry on the cotangent bundle of S^2.

If $E\colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}$ is the Kepler Hamiltonian, the Ligon–Schaaf regularization map is a symplectic embedding

$$\Phi\colon E^{-1}(-\infty, 0) \to T^*S^2 \setminus S^2 \subset T^*\mathbb{R}^3.$$

Here we use the inclusion $T^*S^2 = \{(q,p) \in T^*\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3 \mid |q| = 1, \langle q, p\rangle = 0\}$. Given $(q,p) \in E^{-1}(-\infty, 0)$ abbreviate

$$\begin{aligned} u = u(q,p) &:= \Big(\sqrt{-2E(q,p)}|q|p, |p|^2|q| - 1\Big) \in \mathbb{R}^3 \\ v = v(q,p) &:= \Big(|q|^{-1}q - \langle q,p\rangle p, \sqrt{-2E(q,p)}\langle q,p\rangle\Big) \in \mathbb{R}^3 \\ \phi = \phi(q,p) &:= \sqrt{-2E(q,p)}\langle q,p\rangle \end{aligned}$$

and set

$$\Phi(q,p) = \Big((\cos\phi)u + (\sin\phi)v, \frac{1}{\sqrt{-2E}}\big((\sin\phi)u - (\cos\phi)v\big)\Big).$$

A short computation, worked out in Lemma 4.3.1, shows that $|u| = |v| = 1$ and that $\langle u, v\rangle = 0$: we only need to insert the definition of the Kepler energy $E = \frac{|p|^2}{2} - \frac{1}{|q|}$. We call the map Φ the Ligon–Schaaf embedding. It has some remarkable properties, which we list below as (i)–(iii). Proofs of these properties can be found in [67, 112, 163]. We give alternative proofs of a part of property (i), namely that Φ is symplectic, and of properties (ii) and (iii) in Section 4.3.1.

(i) The map Φ is a symplectic embedding of the space $E^{-1}(-\infty, 0)$ into $T^*S^2 \setminus S^2$ whose image is the complement of the fiber over the north pole in $T^*S^2 \setminus S^2$.

(ii) $E = D \circ \Phi$, where $D\colon T^*S^2 \setminus S^2 \to \mathbb{R}$ is the *Delaunay Hamiltonian* given by

$$D(x,y) = -\frac{1}{2|y|^2}, \quad (x,y) \in T^*S^2 \setminus S^2.$$

Here we think again of a point $x \in S^2$ as a unit vector in $\mathbb{R}^3$ and then identify its cotangent space $T^*_xS^2$ with $x^\perp \subset \mathbb{R}^3$.

Note that the flow of the Delaunay Hamiltonian on each energy level is given by the geodesic flow of S^2 up to constant reparametrization. It follows that the regularization of Ligon–Schaaf embeds the Kepler flow in the geodesic flow on the round sphere, reminiscent of Moser regularization. However, the two regularizations are different. Moser regularization involves a non-constant reparametrization, whereas the regularization of Ligon–Schaaf does not.

The third property of the map Φ is the following. Abbreviate by

$$\mu\colon T^*S^2 \setminus S^2 \to \mathbb{R}^3, \quad (x,y) \mapsto x \times y$$

where $x \times y$ is the cross product of x and y interpreted as vectors in $\mathbb{R}^3$ via $T^*S^2 \subset T^*\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3$. Readers familiar with the theory of moment maps, briefly discussed in Section 3.2 and see for example [172], recognize μ as the moment map of the $SO(3)$ action on T^*S^2 obtained by extending the rotation of the two-sphere to a physical transformation of the cotangent bundle. The third property of the Ligon–Schaaf embedding tells us how Φ interchanges the $SO(3)$ action of T^*S^2 with the symmetries of the Kepler problem.

(iii) If L is the angular momentum and A is the Runge–Lenz vector, then the following holds
$$\left(\frac{(A_2,-A_1)}{\sqrt{-2E}}, L\right) = \mu \circ \Phi.$$

In view of the standard commutation relation in the Lie algebra $so(3)$ of the Lie group $SO(3)$ we obtain the Poisson commutation relations

$$\left\{\frac{A_1}{\sqrt{-2E}}, L\right\} = \frac{A_2}{\sqrt{-2E}}, \quad \left\{L, \frac{A_2}{\sqrt{-2E}}\right\} = \frac{A_1}{\sqrt{-2E}}, \quad \left\{\frac{A_2}{\sqrt{-2E}}, \frac{A_1}{\sqrt{-2E}}\right\} = L.$$

Because E Poisson commutes with A and L, this can be rephrased as

$$\{A_1, L\} = A_2, \quad \{L, A_2\} = A_1, \quad \{A_2, A_1\} = (-2E)L.$$

The reader is cordially invited to check these commutation relations directly. We remark that, as for the Moser regularization, but different from the Levi-Civita regularization, the restriction to the planar case is not necessary and the Ligon–Schaaf regularization can readily be generalized to arbitrary dimensions.

4.3.1 Proof of some of the properties of the Ligon–Schaaf map

We prove these properties by direct computation. First some lemmas.

Lemma 4.3.1. *The vectors u and v have unit length and are orthogonal.*

Proof. This is a direct computation. We check that $|u| = 1$:

$$\begin{aligned}
|u|^2 &= (-2E)|q|^2|p|^2 + |p|^4|q|^2 - 2|p|^2|q| + 1 \\
&= \left(-|p|^2 + \frac{2}{|q|}\right)|q|^2|p|^2 + |p|^4|q|^2 - 2|p|^2|q| + 1 \\
&= -|q|^2|p|^4 + 2|q||p|^2 + |p|^4|q|^2 - 2|p|^2|q| + 1 = 1.
\end{aligned}$$

We check that $|v| = 1$

$$|v|^2 = \frac{|q|^2}{|q|^2} - 2\frac{\langle q,p\rangle^2}{|q|} + |p|^2\langle q,p\rangle + \left(-|p|^2 + \frac{2}{|q|}\right)\langle q,p\rangle^2 = 1.$$

We check that $\langle u, v\rangle = 0$:

$$\langle u, v\rangle = \sqrt{-2E}\left(\langle q, p\rangle - |q||p|^2\langle q, p\rangle + |p|^2|q|\langle q, p\rangle - \langle q, p\rangle\right) = 0. \qquad \square$$

The following lemma implies that Φ is an exact symplectomorphism.

Lemma 4.3.2. *The Ligon–Schaaf embedding pulls back the canonical* 1*-form* $y \cdot dx$ *on* $T^*S^2 \setminus S^2$ *to*

$$\Phi^* y \cdot dx = p \cdot dq - 2d\langle q, p\rangle.$$

In particular, Φ *is a symplectomorphism.*

Proof. From the above we know that $u \cdot du = 0$, $v \cdot dv = 0$, $d(u \cdot v) = u \cdot dv + v \cdot du = 0$. We will use these identities in the following computations.

$$\begin{aligned}
\Phi^*\lambda_c = \Phi^* y \cdot dx &= \frac{1}{\sqrt{-2E}}\left(\sin\phi u - \cos\phi v\right) \cdot d(\cos\phi u + \sin\phi v) \\
&= \frac{1}{\sqrt{-2E}}\left(\sin\phi u - \cos\phi v\right) \cdot \left(\cos\phi du + \sin\phi dv - u\sin\phi d\phi + v\cos\phi d\phi\right) \\
&= \frac{1}{\sqrt{-2E}}\left(\sin^2\phi u \cdot dv - u^2\sin^2\phi d\phi - \cos^2\phi v \cdot du - v^2\cos^2\phi d\phi\right) \\
&= \frac{1}{\sqrt{-2E}}\left(u \cdot dv - d\phi\right).
\end{aligned}$$

We now substitute the expressions for u, v and ϕ into this, and massage the resulting expression,

$$\begin{aligned}
\frac{u \cdot dv}{\sqrt{-2E}} - \frac{1}{\sqrt{-2E}}d\phi &= |q|p \cdot d\left(\frac{q}{|q|} - \langle q, p\rangle p\right) + \left(|p|^2|q| - 1\right) d\langle q, p\rangle \\
&\quad + \left(|p|^2|q| - 1\right)\langle q, p\rangle \frac{d\sqrt{-2E}}{\sqrt{-2E}} - \frac{d\sqrt{-2E}}{\sqrt{-2E}}\langle q, p\rangle - d\langle q, p\rangle \\
&= p \cdot dq + |q|\langle q, p\rangle d\frac{1}{|q|} - |q||p|^2 d\langle q, p\rangle - \langle q, p\rangle |q| p \cdot dp \\
&\quad + (|p|^2|q| - 1)d\langle q, p\rangle + \langle q, p\rangle(|p|^2|q| - 2)\frac{d\sqrt{-2E}}{\sqrt{-2E}} - d\langle q, p\rangle \\
&= p \cdot dq - 2d\langle q, p\rangle - |q|\langle q, p\rangle d\left(\frac{1}{2}|p|^2 - \frac{1}{|q|}\right) \\
&\quad - |q|(-2E)\langle q, p\rangle \frac{d\sqrt{-2E}}{\sqrt{-2E}} \\
&= p \cdot dq - 2d\langle q, p\rangle - |q|\langle q, p\rangle dE - |q|\sqrt{-2E}d\sqrt{-2E}\langle q, p\rangle \\
&= p \cdot dq - 2d\langle q, p\rangle + \langle q, p\rangle\left(-|q|dH + |q|dE\right) \\
&= p \cdot dq - 2d\langle q, p\rangle. \qquad \square
\end{aligned}$$

We now check property (ii) directly with Lemma 4.3.1

$$D \circ \Phi = -\frac{1}{2}\frac{(-2E)}{\cos^2\phi + \sin^2\phi} = E.$$

Proof of property (iii). Note that $x \times y = -\frac{1}{\sqrt{-2E}} u \times v$. The third component is given by

$$(x \times y)_3 = -\frac{1}{\sqrt{-2E}}(u_1 v_2 - u_2 v_1) = -|q|\frac{1}{|q|}(p_1 q_2 - p_2 q_1) = L.$$

The first component of $x \times y$ is given by

$$\begin{aligned}
(x \times y)_1 &= -\frac{1}{\sqrt{-2E}}(u_2 v_3 - u_3 v_2) \\
&= -\frac{1}{\sqrt{-2E}}\left(\sqrt{-2E}|q|p_2 \cdot \sqrt{-2E}\langle q, p\rangle - (|p|^2|q| - 1) \cdot \left(\frac{q_2}{|q|} - \langle q, p\rangle p_2\right)\right) \\
&= -\frac{1}{\sqrt{-2E}}\left((-|p|^2|q| + 2)\langle q, p\rangle p_2 + (|p|^2|q| - 1)\langle q, p\rangle p_2 - |p|^2 q_2 + \frac{q_2}{|q|}\right) \\
&= -\frac{1}{\sqrt{-2E}}\left(-\langle q, p\rangle p_2 + 2\langle q, p\rangle p_2 - |p|^2 q_2 + \frac{q_2}{|q|}\right) \\
&= -\frac{1}{\sqrt{-2E}}\left(q_1 p_1 p_2 - q_2 p_1^2 + \frac{q_2}{|q|}\right) = \frac{A_2}{\sqrt{-2E}}.
\end{aligned}$$

The second component of $x \times y$ is given by

$$\begin{aligned}
(x \times y)_2 &= -\frac{1}{\sqrt{-2E}}(u_3 v_1 - u_1 v_3) \\
&= \frac{1}{\sqrt{-2E}}\left(\sqrt{-2E}|q|p_1 \cdot \sqrt{-2E}\langle q, p\rangle - (|p|^2|q| - 1) \cdot \left(\frac{q_1}{|q|} - \langle q, p\rangle p_1\right)\right) \\
&= \frac{1}{\sqrt{-2E}}\left((-|p|^2|q| + 2)\langle q, p\rangle p_1 + (|p|^2|q| - 1)\langle q, p\rangle p_1 - |p|^2 q_1 + \frac{q_1}{|q|}\right) \\
&= \frac{1}{\sqrt{-2E}}\left(-\langle q, p\rangle p_1 + 2\langle q, p\rangle p_1 - |p|^2 q_1 + \frac{q_1}{|q|}\right) \\
&= \frac{1}{\sqrt{-2E}}\left(p_2(p_1 q_2 - p_2 q_1) + \frac{q_1}{|q|}\right) = -\frac{A_1}{\sqrt{-2E}}.
\end{aligned}$$

□

Chapter 5

The Restricted Three-Body Problem

Celestial mechanics has a long history, see for example the encyclopedic works of Hagihara [109] and Szebehely [231], or the Scholarpedia article on the three-body problem by Chenciner [53] or on Celestial Mechanics by Ferraz-Mello [83] and the references therein. The first basic problem in celestial mechanics going

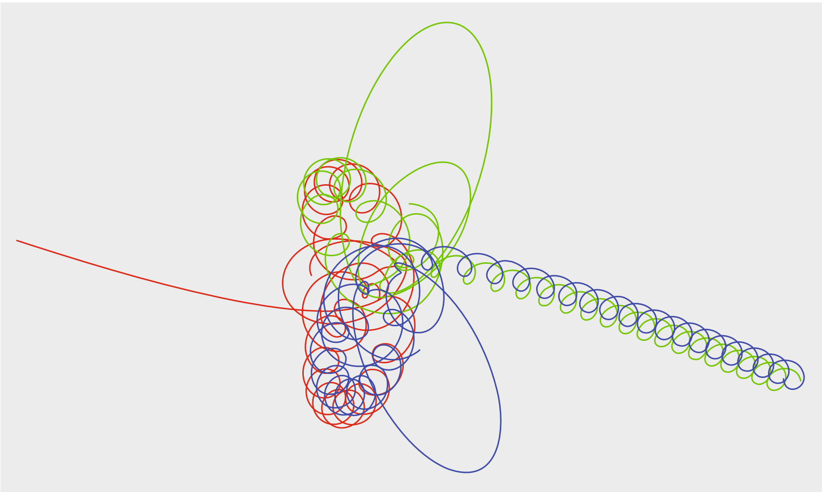


Figure 5.1: Three primaries moving in the unrestricted three-body problem.

beyond the Kepler problem we have seen in Sections 3.3 and 3.5 is of course the three-body problem. Even the planar three-body problem is very complicated due to its high, twelve-dimensional phase space, although this can be reduced by using symmetries, and the presence of three-body collisions, which cannot always be regularized. The Hamiltonian describing this problem is given by

$$H = \sum\nolimits_{i=1}^{3} \frac{1}{2m_i}|p_i|^2 - \sum\nolimits_{i<j} \frac{m_i m_j}{|q_i - q_j|},$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_5

and is a function on $T^*(\mathbb{R}^6 \setminus \Delta)$, where

$$\Delta = \{(q_1, q_2, q_3) \in \mathbb{R}^6 \mid q_i \neq q_j \text{ for } i \neq j\}.$$

Here we think of (q_i, p_i) as a point in $T^*\mathbb{R}^2$. A picture of a sample orbit is given in Figure 5.1.

As stated in the introduction, we will restrict our attention in this monograph to the restricted three-body problem. We will describe this problem in detail in the next section: in short, the restricted three-body problem describes the dynamics of the three-body problem in the limit $m_3 \to 0$. This limit provides both a simplification and a reasonable approximation in many relevant situations: we shall see in the next section what simplifications arise.

The restricted three-body problem is a good approximation to many real-world situations. For example, a satellite is much lighter than the Earth and Sun. Other situations where the restricted three-body problem plays a role is the solar system: both the Sun and Jupiter are much heavier than all the other bodies combined.[3] Even better approximations are Pluto-Charon-Satellite, and Jupiter-Europa-Satellite. We will focus purely on the mathematical aspects.

5.1 The restricted three-body problem in an inertial frame

The first ingredient in the restricted three-body problem are two masses, the *primaries*, which we refer to as the earth and the moon. We rescale the total mass to one so that for some $\mu \in [0, 1]$ the mass of the moon equals μ and the mass of the earth equals $1 - \mu$. Here we allow the mass of the moon to be bigger than the mass of the earth, although in such a situation one might prefer to change the names of the primaries. The earth and the moon move in three-dimensional Euclidean space $\mathbb{R}^3$ according to Newton's law of gravitation and we denote their time-dependent positions by $e(t) \in \mathbb{R}^3$ respectively $m(t) \in \mathbb{R}^3$ for $t \in \mathbb{R}$.

The second ingredient is a massless object referred to as the satellite. Because the satellite is massless it does not influence the movements of the earth and the moon. On the other hand the earth and the moon attract the satellite according to Newton's law of gravitation. The goal of the problem is to get an understanding of the dynamics of the satellite which can be quite intricate. If q denotes the position of the satellite and p its momentum, then the Hamiltonian of the satellite in the inertial system is given according to Newton's law of gravitation by

$$E_t(q, p) = \frac{1}{2}|p|^2 - \frac{\mu}{|q - m(t)|} - \frac{1-\mu}{|q - e(t)|} \tag{5.1}$$

[3]The planets combine to less than 0.4 Jupiter mass, and the Sun is more than one thousand Jupiter masses.

namely the sum of kinetic energy and Newton's potential. We abbreviate this Hamiltonian by E and not by H in order to distinguish it from the Hamiltonian of the restricted three-body problem in rotating coordinates. Note that because the earth and the moon are moving, the Hamiltonian is not autonomous, i.e., it depends on time. Actually, because we have to avoid collisions of the satellite with one of the primaries even the domain of definition of the Hamiltonian is time-dependent, namely

$$E_t \colon T^*\big(\mathbb{R}^3 \setminus \{e(t), m(t)\}\big) \to \mathbb{R}.$$

In particular, because the Hamiltonian depends on time it is not preserved under the flow of its time-dependent Hamiltonian vector field, i.e., preservation of energy does not hold.

If the satellite moves in the ecliptic, i.e., the plane spanned by the orbits of the earth and the moon, after choosing suitable coordinates such that $e(t), m(t) \in \mathbb{R}^2$ for every $t \in \mathbb{R}$, the domain of definition of the Hamiltonian becomes

$$E_t \colon T^*\big(\mathbb{R}^2 \setminus \{e(t), m(t)\}\big) \to \mathbb{R}.$$

This is referred to as the *planar restricted three-body problem*, while if the satellite is allowed to move in three-dimensional space, the problem is called the *spatial restricted three-body problem*. In the following we focus on the planar case. This is due to the fact that the question about global surfaces of section only makes sense in the planar case. A further specialization is obtained by assuming that the earth and moon move on circles about their common center of mass. After choosing suitable coordinates their time-dependent positions are given by

$$e(t) = -\mu\big(\cos(t), -\sin(t)\big), \quad m(t) = (1-\mu)\big(\cos(t), -\sin(t)\big). \tag{5.2}$$

This problem is referred to as the *circular planar restricted three-body problem*. Of course there is also a circular spatial restricted three-body problem. The amazing thing about the circular case is that after a time-dependent transformation which puts the earth and moon at rest, the Hamiltonian of the circular restricted three-body problem in rotating coordinates becomes autonomous, i.e., independent of time. In particular, it is preserved along its flow. This surprising observation is due to Jacobi. We first explain time-dependent transformations.

5.2 Time-dependent transformations

Suppose that (M, ω) is a symplectic manifold and $E \in C^\infty(M \times \mathbb{R}, \mathbb{R})$ and $L \in C^\infty(M \times \mathbb{R}, \mathbb{R})$ are two time-dependent Hamiltonians. For $t \in \mathbb{R}$ abbreviate $E_t = E(\cdot, t) \in C^\infty(M)$ and similarly L_t. This gives rise to two time-dependent Hamiltonian vector fields X_{E_t} and X_{L_t}. For simplicity let us assume that the flows of the Hamiltonian vector fields ϕ_E^t and ϕ_L^t exist for all times. One can consider more complicated situations where the domains of definitions of the two Hamiltonians themselves depend on time. This actually happens in the restricted three-body problem. Nevertheless, the treatment of this more general case does not require new ingredients apart from a notational nightmare.

Define the time-dependent Hamiltonian function

$$L\diamond E \in C^\infty(M \times \mathbb{R}, \mathbb{R})$$

by

$$(L\diamond E)(x,t) = L(x,t) + E((\phi_L^t)^{-1}x, t), \quad x \in M,\ t \in \mathbb{R}. \tag{5.3}$$

We claim that

$$\phi^t_{L\diamond E} = \phi^t_L \circ \phi^t_E, \quad t \in \mathbb{R}. \tag{5.4}$$

To see this, pick $x \in M$. Abbreviate $y = \phi_L^t(\phi_E^t(x))$ and pick further $\xi \in T_yM$. We compute using the fact that ϕ_E^t is symplectic from Theorem 2.2.3

$$\begin{aligned}
\omega\Big(\xi, \frac{d}{dt}(\phi_L^t(\phi_E^t(x))\Big) &= \omega\Big(\xi, X_{L_t}(y) + d\phi_L^t(\phi_E^t(x))X_{E_t}(\phi_E^t(x))\Big)\\
&= dL_t(y)\xi + \omega\Big((d\phi_L^t)^{-1}(y)\xi, X_{E_t}((\phi_L^t)^{-1}(y))\Big)\\
&= dL_t(y)\xi + d(E\circ(\phi_L^t)^{-1})(y)\xi\\
&= d(L\diamond E)_t(y)\xi.
\end{aligned}$$

This establishes (5.4).

Note that even if E and L are autonomous, i.e., independent of time, the Hamiltonian $L\diamond E$ does not need to be autonomous, unless E is invariant under the flow of L.

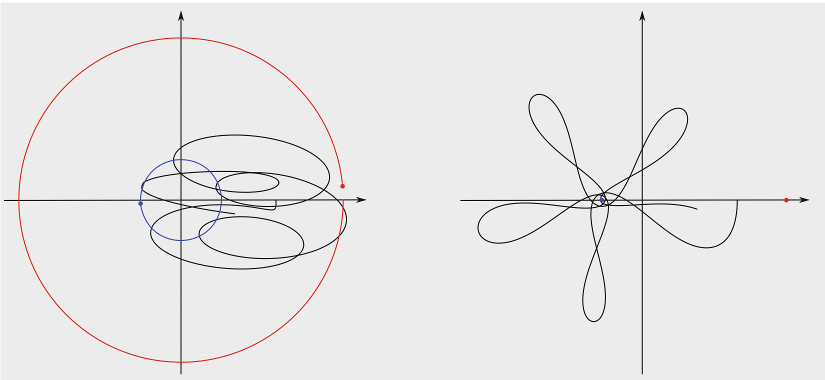


Figure 5.2: An orbit in the unrestricted three-body problem: in sidereal, i.e., non-rotating, coordinates (left) and in rotating coordinates (right).

5.3 The circular restricted three-body problem in a rotating frame

For simplicity we discuss the planar case. The spatial case works analogously. We apply to the Hamiltonian E_t given by (5.1) with positions of the earth and moon determined by (5.2) the time-dependent transformation generated by the angular momentum

$$L \in C^\infty(T^*\mathbb{R}^2, \mathbb{R}), \quad (q,p) \mapsto q_1 p_2 - q_2 p_1.$$

We abbreviate

$$H := L \diamond E.$$

Note that by (2.11) angular momentum generates the rotation. If we abbreviate

$$e = (-\mu, 0), \quad m = (1-\mu, 0),$$

then the Hamiltonian H becomes

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{\mu}{|q-m|} - \frac{1-\mu}{|q-e|} + q_1 p_2 - q_2 p_1. \tag{5.5}$$

Note that this Hamiltonian is *autonomous*. In particular, in the rotating frame the Hamiltonian H is preserved by Theorem 2.2.2. This surprising observation goes back to Jacobi and therefore H is also referred to as the *Jacobi integral*. More precisely, for some historic reasons the integral $-2H$, which of course is preserved under the Hamiltonian flow of H as well, is traditionally called the Jacobi integral.

We point out that the fact that $H = L \diamond E$ is autonomous only holds in the circular case. For example if the primaries move on ellipses with some positive eccentricity, the so-called elliptic restricted three-body problem, the Hamiltonian H does not become time-independent.

Abbreviating the Newtonian potential by

$$V \colon \mathbb{R}^2 \setminus \{e, m\} \to \mathbb{R}, \quad q \mapsto -\frac{\mu}{|q-m|} - \frac{1-\mu}{|q-e|},$$

the Hamiltonian equation of motion becomes

$$\begin{cases} q_1' = p_1 - q_2 \\ q_2' = p_2 + q_1 \\ p_1' = -p_2 - \frac{\partial V}{\partial q_1} \\ p_2' = p_1 - \frac{\partial V}{\partial q_2}. \end{cases} \tag{5.6}$$

For the second derivatives of q we compute

$$q_1'' = p_1' - q_2' = -p_2 - \frac{\partial V}{\partial q_1} - p_2 - q_1 = -2q_2' + q_1 - \frac{\partial V}{\partial q_1}$$

and

$$q_2'' = p_2' + q_1' = p_1 - \frac{\partial V}{\partial q_2} + p_1 - q_2 = 2q_1' + q_2 - \frac{\partial V}{\partial q_2}.$$

Therefore the first-order ODE (5.6) is equivalent to the following second-order ODE

$$\begin{cases} q_1'' = -2q_2' + q_1 - \frac{\partial V}{\partial q_1} \\ q_2'' = 2q_1' + q_2 - \frac{\partial V}{\partial q_2}. \end{cases} \tag{5.7}$$

To give the additional rotational term a physical interpretation we complete the squares and rewrite (5.5) as

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) - \frac{\mu}{|q-m|} - \frac{1-\mu}{|q-e|} - \frac{1}{2}q^2. \tag{5.8}$$

The last three terms only depend on q and we introduce the so-called *effective potential*

$$U \colon \mathbb{R}^2 \setminus \{e,m\} \to \mathbb{R}, \quad q \mapsto -\frac{\mu}{|q-m|} - \frac{1-\mu}{|q-e|} - \frac{1}{2}q^2 = V(q) - \frac{1}{2}q^2.$$

Using this abbreviation the Hamiltonian H can be written more compactly as

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) + U(q). \tag{5.9}$$

The effective potential consists of the Newtonian potential for the earth and the moon plus the additional term $-\frac{1}{2}q^2$. The additional term gives rise to a new force just experienced in rotating coordinates, namely the *centrifugal force*. The Hamiltonian H in (5.9) is not a mechanical Hamiltonian anymore, i.e., it does not just consist of kinetic plus potential energy. Instead of that the Hamiltonian contains a twist in the kinetic part and is therefore a magnetic Hamiltonian as discussed in Section 2.3.4. The twist in the kinetic part can be interpreted in terms of physics as an additional force, namely the *Coriolis force*. Different from the gravitational force and the centrifugal force which only depend on the position of the satellite the Coriolis force depends on its velocity, like the Lorentz force for a particle moving in a magnetic field. This explains why the Hamiltonian of the restricted three-body problem in rotating coordinates becomes a magnetic Hamiltonian. There are now four forces acting on the satellite in the rotating coordinate system, the gravitational force of the earth, the gravitational force of the moon, the centrifugal force, as well as the Coriolis force. This vividly shows that the dynamics of the restricted three-body problem is highly intricate.

5.4 The five Lagrange points

In this section we discuss the critical points of the Hamiltonian H given by (5.9). We immediately observe that the projection map $\pi\colon \mathbb{R}^4 = \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ given by $(q,p) \mapsto q$ induces a bijection

$$\pi|_{\mathrm{crit}(H)} \colon \mathrm{crit}(H) \to \mathrm{crit}(U). \tag{5.10}$$

Indeed, the inverse map for a critical point $(q_1, q_2) \in \mathrm{crit}(U)$ is given by

$$\big(\pi|_{\mathrm{crit}(H)}\big)^{-1}(q_1, q_2) = (q_1, q_2, q_2, -q_1).$$

We explain now that if $\mu \in (0,1)$, the effective potential $U = U_\mu$ has five critical points. The critical points of U are called *Lagrange points*.

Note that U is invariant under reflection at the axis of earth and moon $(q_1, q_2) \mapsto (q_1, -q_2)$. Therefore either a critical point of U lies on the axis of earth and moon, i.e., the fixed point set of the reflection, or it appears in pairs. It turns out that there are three critical points of U on the axis of earth and moon. These *collinear points* were discovered already by Euler and they are saddle points of U. Moreover, there is one pair of non-collinear critical points of U discovered by Lagrange. It turns out that these points build equilateral triangles with the earth and the moon and for this reason they are referred to as *equilateral points*. The equilateral points are maxima of U. A plot of the potential function is sketched in Figure 5.3.

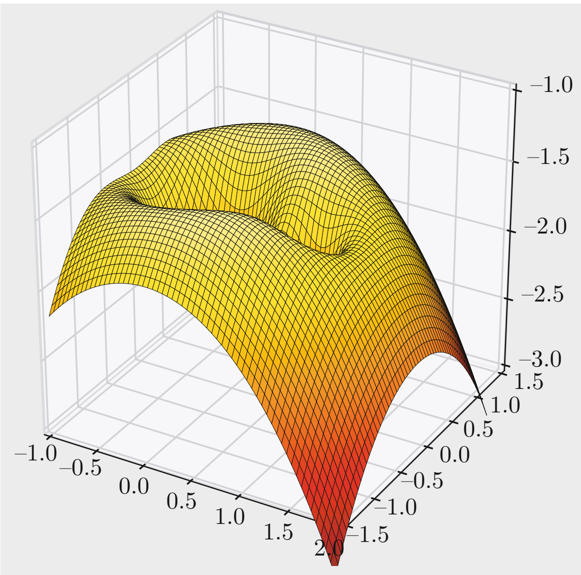


Figure 5.3: The potential function U.

We first discuss the two equilateral Lagrange points following the book by Abraham–Marsden [3]. Because of the reflection symmetry we can restrict our attention to the upper half-space $\mathbb{R} \times (0, \infty)$. Because the distance between earth and moon is one, i.e., $|e - m| = 1$, we have a diffeomorphism

$$\phi \colon \mathbb{R} \times (0, \infty) \to \Theta$$

between the upper half-space and the half-strip

$$\Theta = \{(\rho, \sigma) \in (0, \infty)^2 : \rho + \sigma > 1,\ |\rho - \sigma| < 1\}$$

which is given by

$$\phi(q) = \big(|q - m|, |q - e|\big), \quad q \in \mathbb{R} \times (0, \infty).$$

Consider the smooth function

$$V \colon \Theta \to \mathbb{R}, \quad V := U \circ \phi^{-1}.$$

Critical points of V correspond to critical points of the effective potential U on the upper half-space. To give an explicit description of V in terms of the variables ρ and σ we compute for $q \in \mathbb{R} \times (0, \infty)$

$$\begin{aligned}
|q|^2 &= \mu|q|^2 + (1-\mu)|q|^2 \\
&= \mu\big(\rho^2 + 2\langle m, q\rangle - m^2\big) + (1-\mu)\big(\sigma^2 + 2\langle e, q\rangle - e^2\big) \\
&= \mu\rho^2 + 2\mu(1-\mu)\langle 1, q\rangle - \mu(1-\mu)^2 \\
&\quad + (1-\mu)\sigma^2 - 2\mu(1-\mu)\langle 1, q\rangle - (1-\mu)\mu^2 \\
&= \mu\rho^2 + (1-\mu)\sigma^2 - \mu(1-\mu).
\end{aligned}$$

Therefore V as function of ρ and σ reads

$$V(\rho, \sigma) = -\frac{\mu}{\rho} - \frac{1-\mu}{\sigma} - \frac{1}{2}\big(\mu\rho^2 + (1-\mu)\sigma^2 - \mu(1-\mu)\big). \tag{5.11}$$

Its differential is given by

$$dV(\rho, \sigma) = \frac{\mu(1-\rho^3)}{\rho^2} d\rho + \frac{(1-\mu)(1-\sigma^3)}{\sigma^2} d\sigma.$$

Hence V has a unique critical point at $(1, 1) \in \Theta$. The Hessian at the critical point $(1, 1)$ is given by

$$\mathrm{Hessian}_V(1, 1) = \begin{pmatrix} -3\mu & 0 \\ 0 & -3(1-\mu) \end{pmatrix}.$$

We conclude that $(1, 1)$ is a maximum. Going back to the original coordinates we define the Lagrange point ℓ_4 as

$$\ell_4 = \phi^{-1}(1, 1) := \big(\tfrac{1}{2} - \mu, \tfrac{\sqrt{3}}{2}\big). \tag{5.12}$$

We define the Lagrange point ℓ_5 as the point obtained by reflecting ℓ_4 at the axis through earth and moon

$$\ell_5 := \left(\tfrac{1}{2} - \mu, -\tfrac{\sqrt{3}}{2}\right). \tag{5.13}$$

By reflection symmetry of U the Lagrange point ℓ_5 is also a maximum of U and it is the only critical point in the lower half-space $\mathbb{R} \times (-\infty, 0)$. We summarize what we proved so far

Lemma 5.4.1. *The only critical points of U on $\mathbb{R}^2 \setminus (\mathbb{R} \times \{0\})$, i.e., the complement of the axis through earth and moon are ℓ_4 and ℓ_5 and they are maxima of U.*

We next discuss the collinear critical points. For this purpose we consider the function

$$u := U|_{\mathbb{R}\setminus\{-\mu,1-\mu\}} \colon \mathbb{R} \setminus \{-\mu, 1-\mu\} \to \mathbb{R}, \quad r \mapsto -\frac{\mu}{|r+\mu-1|} - \frac{1-\mu}{|r+\mu|} - \frac{r^2}{2}.$$

Because U is invariant under reflection at the axis of the earth and moon, it follows that critical points of u are critical points of U as well. The second derivative of u is given by

$$u''(r) = -\frac{2\mu}{|r+\mu-1|^3} - \frac{2(1-\mu)}{|r+\mu|^3} - 1 < 0. \tag{5.14}$$

Therefore u is strictly concave. Since at the singularities at $-\mu$ and $1-\mu$ as well as at $-\infty$ and ∞ the function u goes to $-\infty$, we conclude that the function u attains precisely three maxima, one at a point $-\mu < \ell_1 < 1-\mu$, one at a point $\ell_2 > 1-\mu$ and one at a point $\ell_3 < -\mu$. The points ℓ_1, ℓ_2, ℓ_3 are referred to as the three collinear Lagrange points. Although one has closed formulas for the position of the Lagrange points ℓ_4 and ℓ_5, a similar closed formula does not exist for ℓ_1, ℓ_2, or ℓ_3. In fact to obtain the exact position of the three collinear Lagrange points one has to solve quintic equations with coefficients depending on μ, see [3, Chapter 10].

Lemma 5.4.2. *The three collinear Lagrange points are saddle points of the effective potential U.*

Remark 5.4.3. Below we will give a self-contained proof of this lemma, but we also want to point to Conley's paper on low-energy transfer orbits. Namely, Formula 5 from [59] implies this lemma, and gives in addition more dynamical information.

Proof. To prove that the collinear Lagrange points are saddle points one needs to show that

$$\det \begin{pmatrix} \frac{\partial^2 U}{\partial q_1^2}(\ell_i) & \frac{\partial^2 U}{\partial q_1 \partial q_2}(\ell_i) \\ \frac{\partial^2 U}{\partial q_1 \partial q_2}(\ell_i) & \frac{\partial^2 U}{\partial q_2^2}(\ell_i) \end{pmatrix} < 0, \quad 1 \le i \le 3.$$

Because U is invariant under reflection at the q_1-axis and the three collinear Lagrange points are fixed points of this reflection, we conclude that

$$\frac{\partial^2 U}{\partial q_1 \partial q_2}(\ell_i) = 0, \quad 1 \le i \le 3.$$

We have already noted in (5.14) that

$$\frac{\partial^2 U}{\partial q_1^2}(\ell_i) < 0, \quad 1 \leq i \leq 3.$$

Therefore it suffices to check that

$$\frac{\partial^2 U}{\partial q_2^2}(\ell_i) > 0, \quad 1 \leq i \leq 3.$$

We leave this as an exercise to the reader to check. The first one who has proved this simultaneously for all mass ratios μ, seems to be Plummer [204]. If you find this exercise a bit exhausting you are in good company, because Plummer mentions at the end of his article

> "I have thought it worth while to give a definite proof of this fact because on this point Dr. Charlier says: 'That is what I have found through numerical calculations for different values of μ, an algebraic demonstration of this fact seeming to be somewhat complicated.'"

Detailed discussions of this fact might for example be found in [3, Chapter 10] or [231, Chapter 4.6.2].

We finish by giving a topological argument to prove the lemma. In order for this argument to work out, we need to assume that the collinear Lagrange points are non-degenerate in the sense that the kernel of the Hessian at them is trivial. By the discussion above this is equivalent to the assumption that

$$\frac{\partial^2 U}{\partial q_1^2}(\ell_i) \neq 0, \quad 1 \leq i \leq 3.$$

Note that the Euler characteristic of the two fold punctured plane satisfies

$$\chi\big(\mathbb{R}^2 \setminus \{e, m\}\big) = -1.$$

Denote by ν_2 the number of maxima of U, by ν_1 the number of saddle points of U, and by ν_0 the number of minima of U. Because U goes to $-\infty$ at infinity as well as at the singularities e and m, it follows from the Poincaré-Hopf index theorem that

$$\nu_2 - \nu_1 + \nu_0 = \chi\big(\mathbb{R}^2 \setminus \{e, m\}\big) = -1. \tag{5.15}$$

By Lemma 5.4.1 we know that ℓ_4 and ℓ_5 are maxima, so that

$$\nu_2 \geq 2. \tag{5.16}$$

Since the three collinear Lagrange points are maxima of u, the restriction of U to the axis through earth and moon, it follows that they are either saddle points or maxima of U. In particular,

$$\nu_0 = 0 \tag{5.17}$$

and therefore

$$\nu_1 + \nu_2 = 5. \tag{5.18}$$

Combining (5.15), (5.17), and (5.18) we conclude that

$$\nu_2 = 2, \quad \nu_1 = 3.$$

This finishes the proof of the lemma in the non-degenerate case. □

Because of the reflection symmetry the function U attains the same value at ℓ_4 and ℓ_5. In view of Lemma 5.4.2 together with the fact that U goes to $-\infty$ at infinity and the singularities e and m, we conclude that it attains its global maximum at the two equilateral Lagrange points. We state this fact in the following corollary.

Corollary 5.4.4. *The effective potential attains its global maximum precisely at the two equilateral Lagrange points and it holds that*

$$\max U = U(\ell_4) = U(\ell_5) = -\frac{3}{2} + \frac{\mu(\mu-1)}{2}.$$

Proof. To compute the value of $U(\ell_4)$ we get using (5.11)

$$U(\ell_4) = V(1,1) = -\mu - (1-\mu) - \frac{1}{2}\big(\mu + 1 - \mu - \mu(1-\mu)\big) = -\frac{3}{2} + \frac{\mu(\mu-1)}{2}.$$

This finishes the proof of the corollary. □

We next discuss the ordering of the critical values of the saddle points of U.

Lemma 5.4.5. *If $\mu \in \big(0, \frac{1}{2}\big)$, the critical values of the collinear Lagrange points are ordered as follows*

$$U(\ell_1) < U(\ell_2) < U(\ell_3). \tag{5.19}$$

If $\mu = \frac{1}{2}$ we have

$$U(\ell_1) < U(\ell_2) = U(\ell_3). \tag{5.20}$$

Remark 5.4.6. If $\mu \in \big(\frac{1}{2}, 1\big)$ one gets from (5.19) by interchanging the roles of the earth and the moon that

$$U(\ell_1) < U(\ell_3) < U(\ell_2).$$

Proof of Lemma 5.4.5. We follow the exposition given by Kim [149]. We first show $U(\ell_1) < U(\ell_2)$ for $\mu \in (0,1)$. Suppose that $-\mu < q < 1-\mu$. Abbreviate $\rho := 1 - \mu - q > 0$ and set $q' := 1 - \mu + \rho$. In the following, we identify $\mathbb{R}$ with

$\mathbb{R} \times \{0\} \subset \mathbb{R}^2$. We estimate

$$\begin{aligned} U(q') - U(q) &= -\frac{\mu}{\rho} - \frac{1-\mu}{1+\rho} - \frac{1}{2}(1-\mu+\rho)^2 + \frac{\mu}{\rho} + \frac{1-\mu}{1-\rho} + \frac{1}{2}(1-\mu-\rho)^2 \\ &= (1-\mu)\left(\frac{1}{1-\rho} - \frac{1}{1+\rho} - 2\rho\right) \\ &= \frac{2(1-\mu)\rho^3}{1-\rho^2} \\ &> 0. \end{aligned}$$

In particular, by choosing $q = \ell_1$ we get

$$U(\ell_1) < U(\ell_1') \leq U(\ell_2)$$

where for the last inequality we used that ℓ_2 is the maximum of the restriction of U to $(1-\mu, \infty)$.

We next show that for $0 < \mu < \frac{1}{2}$ it holds that $U(\ell_2) < U(\ell_3)$. If $q > 1-\mu$ we estimate

$$\begin{aligned} U(-q) - U(q) &= -\frac{\mu}{1-\mu+q} - \frac{1-\mu}{q-\mu} - \frac{q^2}{2} + \frac{\mu}{q-1+\mu} + \frac{1-\mu}{q+\mu} + \frac{q^2}{2} \\ &= \mu\left(\frac{1}{q-(1-\mu)} - \frac{1}{q+(1-\mu)}\right) + (1-\mu)\left(\frac{1}{q+\mu} - \frac{1}{q-\mu}\right) \\ &= \frac{2\mu(1-\mu)}{q^2-(1-\mu)^2} - \frac{2\mu(1-\mu)}{q^2-\mu^2} \\ &= \frac{2\mu(1-\mu)\big((1-\mu)^2-\mu^2\big)}{(q^2-(1-\mu)^2)(q^2-\mu^2)} \\ &= \frac{2\mu(1-\mu)(1-2\mu)}{(q^2-(1-\mu)^2)(q^2-\mu^2)} \\ &> 0. \end{aligned}$$

We choose now $q = \ell_2$ to obtain

$$U(\ell_2) < U(-\ell_2) \leq U(\ell_3)$$

because ℓ_3 is the maximum of the restriction of U to $(-\infty, -\mu)$.

We finally note that if $\mu = \frac{1}{2}$, the effective potential U is also invariant under reflection at the y-axis $(q_1, q_2) \mapsto (-q_1, q_2)$ and ℓ_2 is mapped to ℓ_3 under reflection at the y-axis. This finishes the proof of the lemma. $\square$

Recall from (5.10) that the projection to position space gives a bijection between critical points of the Hamiltonian H and critical points of the effective potential U. For $i \in \{1,2,3,4,5\}$ abbreviate

$$L_i = \pi|^{-1}_{\mathrm{crit}(H)}(\ell_i) \in \mathrm{crit}(H).$$

If $\ell_i = (q_1^i, q_2^i)$ then $L_i = (q_1^i, q_2^i, -q_2^i, q_1^i)$. Note that

$$H(L_i) = U(\ell_i)$$

and if $\mu(L_i)$ denotes the Morse index of L_i as a critical point of H, i.e., the number of negative eigenvalues of the Hessian of H at L_i, we have

$$\mu(L_i) = \mu(\ell_i).$$

In particular, we proved the following theorem.

Theorem 5.4.7. *For $\mu \in (0,1)$ the Morse indices of the five critical points of H satisfy*

$$\mu(L_1) = \mu(L_2) = \mu(L_3) = 1, \quad \mu(L_4) = \mu(L_5) = 2.$$

If $\mu \in \left(0, \frac{1}{2}\right)$ the critical values of H are ordered as

$$H(L_1) < H(L_2) < H(L_3) < H(L_4) = H(L_5).$$

If $\mu = \frac{1}{2}$, then the critical values satisfy

$$H(L_1) < H(L_2) = H(L_3) < H(L_4) = H(L_5).$$

In Figure 5.4 we have plotted the Lagrange value as a function of μ.

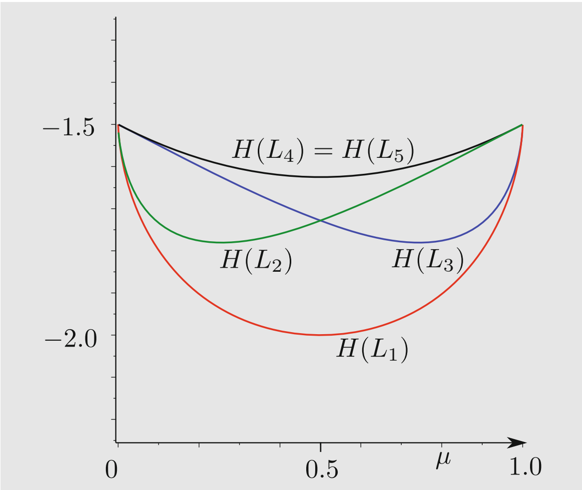


Figure 5.4: Lagrange values as a function of μ.

5.5 Hill's regions

Let H be the Hamiltonian of the planar circular restricted three-body problem in rotating coordinates given by (5.9). Fix $c \in \mathbb{R}$. Because H is autonomous, the energy hypersurface or level set

$$\Sigma_c = H^{-1}(c) \subset T^*\big(\mathbb{R}^2 \setminus \{e, m\}\big)$$

is preserved under the flow of the Hamiltonian vector field of H. Consider the footpoint projection

$$\pi\colon T^*\big(\mathbb{R}^2 \setminus \{e, m\}\big) \to \mathbb{R}^2 \setminus \{e, m\}, \quad (q, p) \mapsto q.$$

The Hill's region of Σ_c is the shadow of Σ_c under the footpoint projection

$$\mathfrak{K}_c := \pi(\Sigma_c) \subset \mathbb{R}^2 \setminus \{e, m\}.$$

Because the first two terms in (5.9) are quadratic and therefore nonnegative we can obtain the Hill's region $\mathfrak{K}_c$ as well as the sublevel set of the effective potential

$$\mathfrak{K}_c = \big\{q \in \mathbb{R}^2 \setminus \{e, m\} : U(q) \leq c\big\}.$$

If the energy lies below the first critical value, i.e., $c < H(L_1)$, the Hill's region

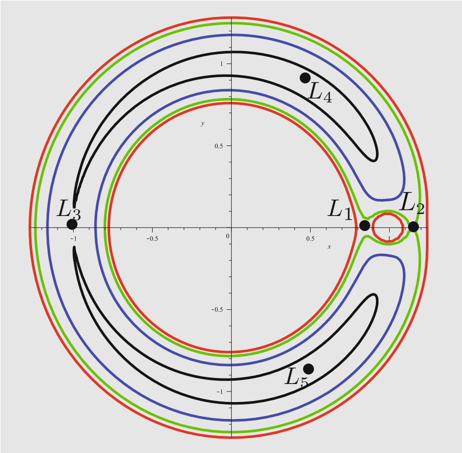


Figure 5.5: The Hill's region for various levels of the Jacobi energy.

has three connected components

$$\mathfrak{K}_c = \mathfrak{K}_c^e \cup \mathfrak{K}_c^m \cup \mathfrak{K}_c^u$$

where the earth e lies in the closure of $\mathfrak{K}_c^e$ and the moon m lies in the closure of $\mathfrak{K}_c^m$. The connected components $\mathfrak{K}_c^e$ and $\mathfrak{K}_c^m$ are bounded, whereas the connected component $\mathfrak{K}_c^u$ is unbounded. Trajectories of the restricted three-body problem in the unbounded component $\mathfrak{K}_c^u$ are referred to as *comets*. Accordingly the energy hypersurface of the restricted three-body problem decomposes into three connected components

$$\Sigma_c = \Sigma_c^e \cup \Sigma_c^m \cup \Sigma_c^u \tag{5.21}$$

where

$$\Sigma_c^e := \{(q,p) \in \Sigma_c,\ q \in \mathfrak{K}_c^e\}$$

and similarly for Σ_c^m and Σ_c^u.

5.6 The rotating Kepler problem

The Hamiltonian of the rotating Kepler problem is given by the Hamiltonian of the restricted three-body problem (5.9) for $\mu = 0$. That means that the moon has zero mass and can be neglected and the satellite is just attracted by the earth like in the Kepler problem. The difference with the usual Kepler problem is that the coordinate system is still rotating. The Hamiltonian of the rotating Kepler problem

$$H \colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}$$

is explicitly given for $(q,p) \in T^*(\mathbb{R}^2 \setminus \{0\})$ by

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{1}{|q|} + q_1 p_2 - q_2 p_1. \tag{5.22}$$

The first two terms are just the Hamiltonian E of the planar Kepler problem while the third term is angular momentum L so that we can write

$$H = E + L.$$

Because E and L Poisson commute, we get

$$\{H, L\} = \{E, L\} + \{L, L\} = 0$$

meaning that H and L Poisson commute as well. In particular, the rotating Kepler problem is an example of a completely integrable system in the sense of Arnold–Liouville, which we discussed in Section 3.4. It is unlikely that the restricted three-body problem is completely integrable for any positive value of μ; in fact, for all but finitely many values of μ analytic integrals can be excluded by work of Poincaré and Xia [205, 244].

If we complete the squares in (5.22), we obtain the magnetic Hamiltonian

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) - \frac{1}{|q|} - \frac{1}{2}|q|^2$$

which we write as

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) + U(q)$$

for the effective potential

$$U\colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$$

given by

$$U(q) = -\frac{1}{|q|} - \frac{1}{2}|q|^2.$$

In contrast to positive μ, this effective potential is rotationally invariant. Therefore its critical set is rotationally invariant as well. We write

$$U(q) = f(|q|)$$

for the function

$$f\colon (0,\infty) \to \mathbb{R}, \quad r \mapsto -\frac{1}{r} - \frac{1}{2}r^2.$$

The differential of f is given by

$$f'(r) = \frac{1}{r^2} - r$$

and therefore f has a unique critical point at $r = 1$ with critical value

$$f(1) = -\frac{3}{2}.$$

We have proved

Lemma 5.6.1. *The effective potential U of the rotating Kepler problem has a unique critical value $-\frac{3}{2}$ and its critical set consists of the circle of radius one around the origin.*

Because critical points of H and U are in bijection via projection as explained in (5.10) and the value of H at a critical point coincides with the value of U of its projection we obtain the following corollary.

Corollary 5.6.2. *The Hamiltonian H of the rotating Kepler problem has a unique critical value $-\frac{3}{2}$.*

5.7 Moser regularization of the restricted three-body problem

In complex notation the Hamiltonian (5.5) of the restricted three-body problem can be written as the map $H\colon T^*(\mathbb{C} \setminus \{-\mu, 1-\mu\}) \to \mathbb{R}$ given by

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{\mu}{|q-1+\mu|} - \frac{1-\mu}{|q+\mu|} + \langle p, iq \rangle.$$

We shift coordinates and put the origin of our coordinate system to the moon to obtain the Hamiltonian $H_m\colon T^*(\mathbb{C}\setminus\{0,-1\})\to\mathbb{R}$ defined by

$$\begin{aligned} H_m(q,p) &= H(q+1-\mu, p-i+i\mu)+\frac{(1-\mu)^2}{2} \\ &= \frac{1}{2}|p|^2-\frac{\mu}{|q|}-\frac{1-\mu}{|q+1|}+\langle p, iq\rangle+\langle q,\mu-1\rangle \\ &= \frac{1}{2}|p|^2-\frac{\mu}{|q|}+\langle p, iq\rangle-\frac{1-\mu}{|q+1|}-(1-\mu)q_1. \end{aligned} \tag{5.23}$$

For $c\in\mathbb{R}$ we consider the energy hypersurface $\Sigma_c=H_m^{-1}(c)$. We switch the roles of q and p and think of p as the base coordinate and q as the fiber coordinate. The energy hypersurface is then a subset $\Sigma_c\subset T^*\mathbb{C}\subset T^*S^2$ where for the last inclusion we think of the sphere as $S^2=\mathbb{C}\cup\{\infty\}$ via stereographic projection.

We examine if the closure $\overline{\Sigma_c}$ is regular in the fiber above ∞. For this purpose we examine how the terms in (5.23) transform under chart transition. The chart transition from the chart given by stereographic projection at the north pole to the chart given by stereographic projection at the south pole is given by

$$\phi\colon \mathbb{C}\setminus\{0\}\to\mathbb{C}\setminus\{0\},\quad p\mapsto\frac{1}{p}=\frac{\overline{p}}{|p|^2}$$

where $\overline{p}=p_1-ip_2$ denotes the complex conjugate of p. In real notation this corresponds to the map

$$(p_1,p_2)\mapsto\left(\frac{p_1}{p_1^2+p_2^2},\frac{-p_2}{p_1^2+p_2^2}\right).$$

We compute the Jacobian of ϕ at $p\in\mathbb{R}^2\setminus\{0\}$ and find

$$d\phi(p)=\frac{1}{(p_1^2+p_2^2)^2}\begin{pmatrix} p_2^2-p_1^2 & -2p_1p_2 \\ 2p_1p_2 & p_2^2-p_1^2\end{pmatrix}.$$

Its determinant is given by

$$\det\big(d\phi(p)\big)=\frac{1}{(p_1^2+p_2^2)^2}.$$

Therefore the inverse transpose of $d\phi(p)$ reads

$$\big(d\phi(p)^{-1}\big)^T=\begin{pmatrix} p_2^2-p_1^2 & -2p_1p_2 \\ 2p_1p_2 & p_2^2-p_1^2\end{pmatrix}.$$

Consequently the exact symplectomorphism

$$d_*\phi\colon T^*(\mathbb{R}^2\setminus\{0\})\to T^*(\mathbb{R}^2\setminus\{0\})$$

is given by

$$d_*\phi(p,q) = \left(\frac{p_1}{p_1^2+p_2^2}, \frac{-p_2}{p_1^2+p_2^2}, (p_2^2-p_1^2)q_1 - 2p_1p_2q_2, 2p_1p_2q_1 + (p_2^2-p_1^2)q_2\right). \quad (5.24)$$

Note that because $\phi^{-1} = \phi$, it holds that

$$(d_*\phi)^{-1} = d_*\phi^{-1} = d_*\phi.$$

Therefore the push forward of kinetic energy to the chart centered at the south pole is given by

$$(d_*\phi)_*\left(\frac{|p|^2}{2}\right) = (d_*\phi)^*\left(\frac{|p|^2}{2}\right) = \frac{1}{2|p|^2}.$$

The push forward of Newton's potential is

$$(d_*\phi)_*\left(\frac{1}{|q|}\right) = \frac{1}{|p|^2|q|}.$$

The angular momentum $L = q_1p_2 - q_2p_1$ transforms as

$$(d\phi_*)_*L = -L.$$

Assume that $\Omega \subset \mathbb{R}^2$ is an open subset containing the origin 0 and $V\colon \Omega \to \mathbb{R}$ is a smooth function. Set

$$F_V\colon \mathbb{R}^2 \times \Omega \to \mathbb{R}, \quad (p,q) \mapsto V(q)$$

and define

$$F^V : (d_*\phi)\big((\mathbb{R}^2 \setminus \{0\}) \times \Omega\big) \cup \big(\{0\} \times \mathbb{R}^2\big) \to \mathbb{R}$$

by

$$F^V(p,q) = \begin{cases} F_V\big(d_*\phi(p,q)\big) & p \neq 0 \\ V(0) & p = 0. \end{cases}$$

It follows from the transformation formula (5.24) that F^V is smooth.

We now choose $\Omega = \mathbb{C} \setminus \{-1\}$ and

$$V\colon \Omega \to \mathbb{R}, \quad q \mapsto -\frac{1-\mu}{|q+1|} - (1-\mu)q_1.$$

It follows that the Hamiltonian H_m in the chart given by stereographic projection at the south pole is given by the map

$$(q,p) \mapsto \frac{1}{2|p|^2} - \frac{1}{|p|^2|q|} - L(q,p) - F^V(p,q).$$

Because F^V extends smoothly to $p = 0$, the energy hypersurface Σ_c extends smoothly to $p = 0$ as well and the intersection of its closure with the fiber over 0 is given by

$$\left\{\frac{1}{|q|} = \frac{1}{2} : q \in \mathbb{R}^2\right\} = \big\{|q| = 2 : q \in \mathbb{R}^2\big\},$$

i.e., a circle of radius 2.

The origin in the chart obtained by stereographic projection at the south pole corresponds to the north pole, respectively the point $\infty \in S^2$. The fact that we obtained a circle as the intersection, depends on our specific choice of coordinates on the sphere S^2. However, because convexity is preserved under linear transformations we proved the following lemma.

Lemma 5.7.1. *For every $c \in \mathbb{R}$ the closure of the energy hypersurface $\Sigma_c = H_m^{-1}(c) \subset T^*S^2$ is regular in the fiber over ∞ and $\overline{\Sigma}_c \cap T^*_\infty S^2$ is convex.*

Recall from (5.21) that if the energy c is less then the first critical value $H(L_1)$ the energy hypersurface Σ_c has three connected components Σ_c^e, Σ_c^m, Σ_c^u, the first close to the earth, the second close to the moon and the third consisting of comets. The regularization described above compactifies the component Σ_c^m to $\overline{\Sigma}_c^m$. Of course the same can be done with the component around the earth to obtain the regularized component $\overline{\Sigma}_c^e$. Indeed, the roles of the earth and moon can always be interchanged simply by replacing μ by $1-\mu$. In the following we discuss the regularized component $\overline{\Sigma}_c^e$. But by interchanging the roles of the earth and the moon everything said below is also true for the moon. In [9] the following theorem is proved.

Theorem 5.7.2 (Albers–Frauenfelder–Paternain–van Koert). *For energy values below the first critical value, i.e., for $c < H(L_1)$, the regularized energy hypersurface $\overline{\Sigma}_c^e \subset T^*S^2$ is fiberwise star-shaped.*

As an immediate corollary of this theorem we have

Corollary 5.7.3. *Under the assumption of Theorem* 5.7.2 *the restriction of the Liouville one-form λ on T^*S^2 gives a contact form on $\overline{\Sigma}_c^e$. In particular, after reparametrization the regularized flow of the restricted three-body problem around the earth below the first critical value can be interpreted as a Reeb flow.*

We have seen that in the fiber over ∞ the regularized energy hypersurface bounds actually a convex domain. Therefore we ask the following question.

Question 5.7.4. *Under the assumptions of Theorem* 5.7.2 *is $\overline{\Sigma}_c^e$ fiberwise convex in T^*S^2, i.e., after reparametrization can the regularized flow of the restricted three-body problem around the earth below the first critical value be interpreted as a Finsler flow?*

In [158] Lee proved that below the first critical value Hill's lunar problem is fiberwise convex. Hill's lunar problem is a limit problem of the restricted three-body problem where the mass ratio of the earth and moon diverges to ∞ and the satellite moves in a tiny neighborhood of the moon. We discuss this problem in Section 5.8. It is also known that below the first critical value the regularized rotating Kepler problem is fiberwise convex, see [56].

As a further corollary of Theorem 5.7.2 we obtain

Corollary 5.7.5. *Under the assumption of Theorem* 5.7.2 *the regularized energy hypersurface $\overline{\Sigma}_c^e$ is diffeomorphic to $\mathbb{R}P^3$.*

We finish this section by explaining how this corollary follows in a more elementary way from the Fibration theorem of Ehresmann without referring to Theorem 5.7.2. We first recall the following theorem of Ehresmann [76]. A proof can be found for example in [44, Theorem 8.12]

Theorem 5.7.6 (Fibration theorem of Ehresmann)**.** *Assume that $f\colon Y \to X$ is a proper submersion of differential manifolds, then f is a locally trivial fibration.*

As a corollary of Ehresmann's fibration theorem we have

Corollary 5.7.7. *Assume that $F\colon M \times [0,1] \to \mathbb{R}$ is a smooth function such that $F^{-1}(0)$ is compact and for every $r \in [0,1]$ it holds that 0 is a regular value of $F_r := F(\cdot, r)\colon M \to \mathbb{R}$. Then $F_0^{-1}(0)$ and $F_1^{-1}(0)$ are diffeomorphic closed manifolds.*

Proof. That $F_r^{-1}(0)$ is a closed manifold for every $r \in [0,1]$ follows from the assumptions that 0 is a regular value of F_r and $F^{-1}(0)$ is compact. It remains to prove that all these manifolds are diffeomorphic. We show this by applying Theorem 5.7.6 to the projection map

$$\pi\colon Y := F^{-1}(0) \to [0,1], \quad (x,r) \mapsto r.$$

Because Y is compact by assumption, the projection map is proper. It remains to check that π is submersive. If $(x,r) \in Y$, the differential of π is the linear map

$$d\pi_{(x,r)}\colon T_{(x,r)}Y \to \mathbb{R}, \quad (\widehat{x}, \widehat{r}) \mapsto \widehat{r}.$$

The tangent space of Y is given by

$$T_{(x,r)}Y = \big\{(\widehat{x}, \widehat{r}) \in T_xM \times \mathbb{R} : dF_{(x,r)}(\widehat{x}, \widehat{r}) = 0\big\}.$$

Pick $\widehat{r} \in \mathbb{R}$. Because 0 is a regular value of F_r, there exists $\widehat{x} \in T_xM$ such that

$$dF_{(x,r)}(\widehat{x}, 0) = dF_r(\widehat{x}) = -dF_{(x,r)}(0, \widehat{r}).$$

This implies that

$$dF_{(x,r)}(\widehat{x}, \widehat{r}) = dF_{(x,r)}(\widehat{x}, 0) + dF_{(x,r)}(0, \widehat{r}) = 0$$

and therefore $(\widehat{x}, \widehat{r}) \in T_{(x,r)}Y$. Since $\widehat{r} \in \mathbb{R}$ was arbitrary, this shows that $d\pi_{(x,r)}$ is surjective and π is a proper submersion. Now the assertion of the corollary follows from Theorem 5.7.6. □

We now use Corollary 5.7.7 to give a direct proof of Corollary 5.7.5 with no reference to Theorem 5.7.2.

Proof of Corollary 5.7.5. In view of Corollary 5.7.7 the Corollary 5.7.5 can now be proved by a deformation argument. Namely we first switch off the moon to end up in the rotating Kepler problem and then we switch off the rotation as well. To make the dependence of the regularized energy hypersurface on the mass of the moon μ visible we write $\overline{\Sigma}^e_{c,\mu}$. For given $\mu_1 \in (0,1)$ and $c_1 < H_{\mu_1}(L_{1,\mu_1})$ we choose a smooth path $c\colon [0,\mu_1] \to \mathbb{R}$ with the property that $c(\mu) < H_\mu(L_{1,\mu})$ and $c(\mu_1) = c_1$. In view of Corollary 5.7.7 the manifold $\overline{\Sigma}^e_{c_1,\mu_1}$ is diffeomorphic to $\overline{\Sigma}^e_{c(0),0}$ but the last one is just the regularized energy hypersurface of the rotating Kepler problem below the first critical value. A further homotopy which switches off the rotation in the rotating Kepler problem shows that the latter one is diffeomorphic to the regularized energy hypersurface of the (non-rotating) Kepler problem for a negative energy value which by Moser is diffeomorphic to $\mathbb{R}P^3$. This proves the corollary. □

5.8 Hill's lunar problem

5.8.1 Derivation of Hill's lunar problem

While the restricted three-body problem considers the case where the two primaries have comparable masses, Hill's lunar problem [117] deals with the case where the first primary compared to the second one has a much much bigger mass, the second primary has a much much bigger mass than the satellite and the satellite moves very close to the second primary. We show how to derive the Hamiltonian of Hill's lunar problem from the restricted three-body problem by blowing up to coordinates close to the second primary, compare also [176]. Recall from (5.23) that after shifting position and momenta of the Hamiltonian of the restricted three-body problem in order to put the origin of our coordinate system to the moon we obtain the Hamiltonian

$$H_m(q,p) = \frac{1}{2}|p|^2 - \frac{\mu}{|q|} - (1-\mu)\left(\frac{1}{\sqrt{(q_1+1)^2+q_2^2}} + q_1\right) + q_1p_2 - q_2p_1.$$

The diffeomorphism

$$\phi_\mu\colon T^*\mathbb{R}^2 \to T^*\mathbb{R}^2, \quad (q,p) \mapsto (\mu^{\frac{1}{3}}q, \mu^{\frac{1}{3}}p)$$

is conformally symplectic with constant conformal factor $\mu^{\frac{2}{3}}$, i.e.,

$$\phi_\mu^*\omega = \mu^{\frac{2}{3}}\omega$$

for the standard symplectic form on $T^*\mathbb{R}^2$. Define

$$H^\mu\colon T^*(\mathbb{R}^2 \setminus \{(0,0),(-\mu^{-\frac{1}{3}},0)\}) \to \mathbb{R}, \qquad H^\mu := \mu^{-\frac{2}{3}}(H_m \circ \phi_\mu + 1 - \mu).$$

Because ϕ is symplectically conformal with conformal factor $\mu^{\frac{2}{3}}$, it follows that

$$X_{H^\mu} = \phi_\mu^* X_{H_m}.$$

Explicitly, the Hamiltonian H^μ is given by

$$H^\mu(q,p) = \frac{1}{2}|p|^2 - \frac{1}{|q|} - \frac{1}{\mu^{\frac{2}{3}}}\left(\frac{1}{\sqrt{1+2\mu^{\frac{1}{3}}q_1 + \mu^{\frac{2}{3}}|q|^2}} + \mu^{\frac{1}{3}}q_1 - 1\right)$$
$$+ q_1p_2 - q_2p_1 + \frac{\mu^{\frac{1}{3}}}{\sqrt{1+2\mu^{\frac{1}{3}}q_1 + \mu^{\frac{2}{3}}|q|^2}} + \mu^{\frac{2}{3}}q_1.$$

By applying Taylor's formula

$$\frac{1}{\sqrt{1+r}} = 1 - \frac{r}{2} + \frac{3r^2}{8} + \mathcal{O}(r^3)$$

we obtain the following expansion of H^μ in powers of $\mu^{\frac{1}{3}}$

$$H^\mu(q,p) = \frac{1}{2}|p|^2 - \frac{1}{|q|} + q_1p_2 - q_2p_1 - q_1^2 + \frac{1}{2}q_2^2 + \mathcal{O}(\mu^{\frac{1}{3}}).$$

Hence we deduce that as μ goes to zero, the Hamiltonian H^μ converges uniformly in the C^∞-topology on each compact subset to the Hamiltonian $H\colon T^*(\mathbb{R}^2\setminus\{0\}) \to \mathbb{R}$ given by

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{1}{|q|} + q_1p_2 - q_2p_1 - q_1^2 + \frac{1}{2}q_2^2. \tag{5.25}$$

We refer to the Hamiltonian H as *Hill's lunar Hamiltonian.*

The Hamiltonian equation corresponding to Hill's lunar Hamiltonian is the following first-order ODE

$$\begin{cases} q_1' = p_1 - q_2 \\ q_2' = p_2 + q_1 \\ p_1' = -p_2 + 2q_1 - \frac{q_1}{|q|^3} \\ p_2' = p_1 - q_2 - \frac{q_2}{|q|^3} \end{cases} \tag{5.26}$$

which is equivalent to the following second-order ODE

$$\begin{cases} q_1'' = -2q_2' + 3q_1 - \frac{q_1}{|q|^3} \\ q_2'' = 2q_1' - \frac{q_2}{|q|^3}. \end{cases} \tag{5.27}$$

5.8.2 Hill's lunar Hamiltonian

Apart from its rather simple form which makes Hill's lunar Hamiltonian an important testing ground for numerical investigations, see for example [222, 229], a nice feature of it is that it is invariant under two commuting anti-symplectic involutions. Namely $\rho_1, \rho_2\colon T^*\mathbb{R}^2 \to T^*\mathbb{R}^2$ given for $(q,p) \in T^*\mathbb{R}^2$ by the formula

$$\rho_1(q_1,q_2,p_1,p_2) = (q_1,-q_2,-p_1,p_2), \quad \rho_2(q_1,q_2,p_1,p_2) = (-q_1,q_2,p_1,-p_2)$$

are both anti-symplectic involutions such that

$$H \circ \rho_1 = H, \quad H \circ \rho_2 = H.$$

The two anti-symplectic involutions commute and their product is the symplectic involution

$$\rho_1 \circ \rho_2 = \rho_2 \circ \rho_1 = -\mathrm{id}.$$

In contrast to Hill's lunar Hamiltonian, the Hamiltonian of the restricted three-body problem is only invariant under ρ_1 but not under ρ_2.

By completing the squares we can write Hill's lunar Hamiltonian (5.25) in the equivalent form

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) - \frac{1}{|q|} - \frac{3}{2}q_1^2.$$

If one introduces the *effective potential* $U \colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ by

$$U(q) = -\frac{1}{|q|} - \frac{3}{2}q_1^2$$

the Hamiltonian for Hill's lunar problem can be written as

$$H(q,p) = \frac{1}{2}\big((p_1 - q_2)^2 + (p_2 + q_1)^2\big) + U(q).$$

Some level sets of the effective potential are sketched in Figure 5.6. These bound the Hill's regions for the corresponding energy levels.

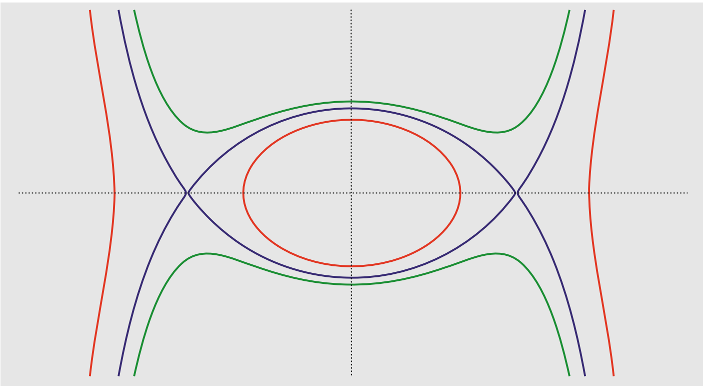


Figure 5.6: The Hill's region of the Hill's lunar problem for various energy levels.

As in the restricted three-body problem there is a one-to-one correspondence between critical points of H and critical points of U. Namely, the footpoint projection

$$\pi \colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}^2 \setminus \{0\}, \quad (q,p) \mapsto q$$

induces a bijection

$$\pi|_{\mathrm{crit}(H)}\colon \mathrm{crit}(H) \to \mathrm{crit}(U).$$

The partial derivatives of U compute to be

$$\frac{\partial U}{\partial q_1} = \frac{q_1}{|q|^3} - 3q_1 = q_1\left(\frac{1}{|q|^3} - 3\right), \qquad \frac{\partial U}{\partial q_2} = \frac{q_2}{|q|^3}. \tag{5.28}$$

The latter implies that at a critical point q_2 has to vanish. Since U has a singularity at the origin, we conclude that at a critical point q_1 does not vanish and therefore the critical set of U is given by

$$\mathrm{crit}(U) = \big\{(3^{-\frac{1}{3}}, 0), (-3^{-\frac{1}{3}}, 0)\big\}.$$

The two critical points of U can be thought of as the limits of the first and second Lagrange point ℓ_1 and ℓ_2 of the restricted three-body problem under the blow up at the moon. Because the remaining Lagrange points ℓ_3, ℓ_4, and ℓ_5 are too far away from the moon, they are not visible in Hill's lunar problem anymore. In particular, the critical set of H reads

$$\mathrm{crit}(H) = \big\{(3^{-\frac{1}{3}}, 0, 0, -3^{-\frac{1}{3}}), (-3^{-\frac{1}{3}}, 0, 0, 3^{-\frac{1}{3}})\big\}.$$

Note that the two critical points of H are fixed by the anti-symplectic involution ρ_1 but interchanged through the anti-symplectic involution ρ_2. Because H is invariant under ρ_2 it attains the same value at the two critical points. That means that Hill's lunar problem has just one critical value which computes to be

$$\begin{aligned} H(3^{-\frac{1}{3}}, 0, 0, -3^{-\frac{1}{3}}) &= U(3^{-\frac{1}{3}}, 0) \\ &= -3^{\frac{1}{3}} - \frac{3}{2}3^{-\frac{2}{3}} \\ &= -3^{\frac{1}{3}} - \frac{3^{\frac{1}{3}}}{2} \\ &= -\frac{3^{\frac{1}{3}}(2+1)}{2} \\ &= -\frac{3^{\frac{4}{3}}}{2}. \end{aligned}$$

To obtain the Hessian of U at its critical points we compute from (5.28)

$$\frac{\partial^2 U}{\partial q_1^2}(\pm 3^{-\frac{1}{3}}, 0) = -\frac{3q_1^2}{|q|^5}\bigg|_{(q_1,q_2)=(\pm 3^{-\frac{1}{3}},0)} = -\frac{3}{|q_1|^3}\bigg|_{q_1=\pm 3^{-\frac{1}{3}}} = -9,$$

$$\frac{\partial^2 U}{\partial q_1 \partial q_2}(\pm 3^{-\frac{1}{3}}, 0) = \frac{\partial^2 U}{\partial q_2 \partial q_1}(\pm 3^{-\frac{1}{3}}, 0) = 0,$$

and finally

$$\frac{\partial^2 U}{\partial q_2^2}(\pm 3^{-\frac{1}{3}}, 0) = \frac{1}{|q|^3}\bigg|_{(q_1,q_2)=(\pm 3^{-\frac{1}{3}},0)} = 3.$$

Therefore the Hessian of U at its critical points is given by

$$\mathrm{Hessian}_U(\pm 3^{-\frac{1}{3}}, 0) = \begin{pmatrix} -9 & 0 \\ 0 & 3 \end{pmatrix}. \tag{5.29}$$

We conclude that the critical points of U are saddle points. It follows that the two critical points of H have Morse index equal to one. We summarize this fact in the following lemma.

Lemma 5.8.1. *Hill's lunar Hamiltonian has a unique critical value at energy $-\frac{3^{\frac{4}{3}}}{2}$. At the critical value it has two critical points both of Morse index one.*

If $c \in \mathbb{R}$ we abbreviate by

$$\Sigma_c = H^{-1}(c)$$

the three-dimensional energy hypersurface of H in the four-dimensional phase space $T^*(\mathbb{R}^2 \setminus \{0\})$. The *Hill's region* for the energy value c is defined as

$$\mathfrak{K}_c = \pi(\Sigma_c) = \big\{ q \in \mathbb{R}^2 \setminus \{0\} : U(q) \leq c \big\}. \tag{5.30}$$

If $c < -\frac{3^{\frac{4}{3}}}{2}$, then the Hill's region $\mathfrak{K}_c$ has three connected components, one bounded and the other two unbounded. We denote the bounded component of $\mathfrak{K}_c$ by $\mathfrak{K}_c^b$ and abbreviate

$$\Sigma_c^b := \big\{ (q,p) \in \Sigma_c : q \in \mathfrak{K}_c^b \big\}. \tag{5.31}$$

5.9 Euler's problem of two fixed centers

For $\mu \in [0,1]$ the Hamiltonian for the (planar) Euler problem of two fixed centers is the function

$$H \colon T^*(\mathbb{C} \setminus \{0,1\}) \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|^2 - \frac{1-\mu}{|q|} - \frac{\mu}{|q-1|}.$$

Its Hamiltonian flow describes the behavior of a massless particle attracted by two primaries of mass μ and $1-\mu$ located at the origin, and at $1 \in \mathbb{C}$, respectively. In contrast to the restricted three-body problem, the two primaries do not rotate around each other. Therefore neither the centrifugal force nor the Coriolis force are present in this system. Mathematically this just amounts to cancelling the rotational term in the restricted three-body problem. If $\mu = \frac{1}{2}$ and is interpreted as a charge instead of a mass, then Euler's problem of two fixed centers can be interpreted as describing the movement of an electron attracted by two protons. This is the situation one faces in the study of a hydrogen molecule. For this reason Euler's problem of two fixed centers was used by Pauli in his quantum theoretical treatment of the hydrogen molecule [201].

A nontrivial feature of Euler's problem of two fixed centers is that it is completely integrable. This is far from obvious since the integral of the problem leads to a nonphysical symmetry. In fact, unless the mass of one of the particles vanishes, the problem is not invariant under rotational symmetry. The integral is the following complicated expression

$$B(q,p) = B_\mu(q,p) = -L(q,p)^2 - L(q,p)p_2 - \frac{(1-\mu)q_1}{|q|} - \frac{\mu(1-q_1)}{|q-1|}$$

where $L(q,p) = p_1 q_2 - p_2 q_1$ is the negative angular momentum. The suspicious reader can check that

$$\{H, B\} = 0. \tag{5.32}$$

For $\mu = 0$ the integral becomes

$$B_0 = -L^2 + A_1$$

where A_1 is the first component of the Runge–Lenz vector. Because both L and A_1 commute with the Kepler Hamiltonian, this establishes (5.32) for the special case $\mu = 0$. The integral was discovered by Euler [78, 79]. We refer to [194] for more information on the history of this problem. In his lectures on dynamics [140] Jacobi explained how the integral can be derived from the method now referred to as Hamilton–Jacobi method by using elliptic coordinates, see also [16]. Thanks to the integral, trajectories of the Hamiltonian flow of H can be expressed in terms of elliptic functions, see [194, 236]. Although it is very unlikely that the restricted

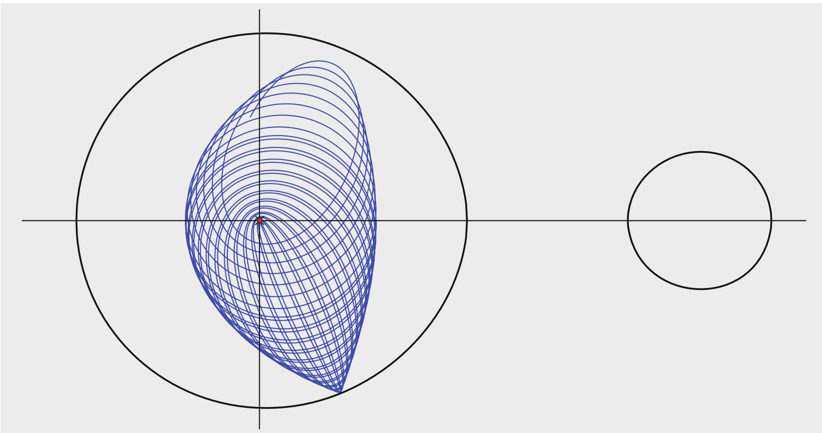


Figure 5.7: An orbit in Euler's problem of two fixed centers.

three-body problem is completely integrable and at least analytic integrals can be ruled out [205, 244], we discuss now two ways how the restricted three-body problem can be homotoped to a completely integrable system. The first homotopy is switching off the moon. This deforms the restricted three-body problem to the rotating Kepler problem which is completely integrable with integral angular momentum. The second homotopy is switching off the rotational term. This homotopes the restricted three-body problem to Euler's problem of two fixed centers. To phrase this homotopy in more physical terms, we think of the earth and moon not just as having masses but also electric charges. The satellite has just mass and is uncharged. If the charges of both the moon and earth are either positive or negative they repel each other according to Coulomb's law and attract each other according to Newton's law of gravitation. If the repulsive Coulomb force has the same magnitude as the attractive gravitational force, then their effect cancels and we are left with Euler's problem of two fixed centers. On the other hand if there is no charge we have the restricted three-body problem. In view of this interpretation we can think of switching off the rotational term of the Hamiltonian of the restricted three-body problem as switching on a charge on the earth and moon. Hence for $\mu \in [0,1]$ and $a \in [0,1]$ we refer to the Hamiltonian

$$H\colon T^*\mathbb{C}\setminus\{-\mu, 1-\mu\} \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|^2 - \frac{\mu}{|q+\mu|} - \frac{1-\mu}{|q-1+\mu|} + a(p_1q_2 - p_2q_1)$$

as the Hamiltonian of the *charged restricted three-body problem*. If $a = 1$ we recover the Hamiltonian of the restricted three-body problem and if $a = 0$ we obtain the Hamiltonian of Euler's problem of two fixed centers after translating the base coordinates by $-\mu$.

The fact that the restricted three-body problem can be homotoped to the completely integrable rotating Kepler problem has important applications to our understanding of the dynamics of the restricted three-body problem since it leads to the orbits of first and second kind discussed in Chapter 7. It should be possible to carry out similar studies with Euler's problem of two fixed centers. Namely periodic orbits of Euler's problem of two fixed centers should give rise to periodic orbits in the charged restricted three-body problem which one can then try to homotope to periodic orbits of the actual restricted three-body problem by switching off the charge. This idea is very old. In fact, in his lectures on the mechanics of the sky [51], Charlier introduced Euler's problem of two fixed centers first in order to come from it to the restricted three-body problem. To our knowledge these ideas have not yet been pursued in practice. However, in view of the speed of modern computers these visions become quite feasible these days.

A beautiful derivation without computation of the complete integrability of Euler's problem was recently given by Albouy [11]. Using an idea of Appell [14] he considers the central or gnomonic projection from the sphere to the plane. Comparing the Euler problem on the plane and on the sphere he discovers that the central projection maps the energy of the Euler problem on the sphere to an integral for the Euler problem on the plane. A nice observation of Albouy is that using the central projection he obtains the integrability of both Euler problems, namely the one on the plane and on the sphere. That the Euler problem on a space of constant curvature like the sphere remains integrable was first noticed by Killing [148].

Lagrange observed [157] that the problem of two fixed centers remains integrable if one adds an elastic force acting from the midpoint of the two masses. In case the two masses are equal the elastic force can be interpreted as the centrifugal force. In particular, in the case of two equal masses one only needs to cancel the Coriolis force in the restricted three-body problem to get a completely integrable system. In other words, the Hamiltonian

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{1}{2\left|q - \frac{1}{2}\right|} - \frac{1}{2\left|q + \frac{1}{2}\right|} - \frac{1}{2}|q|^2$$

admits an integral. We refer to the paper by Hiltebeitel [118] for a comprehensive treatment of which forces one can add to the problem of two fixed centers while still keeping the problem completely integrable.

In [11] Albouy explains that Appell's central projection can also be used to prove the integrability of the Lagrange problem on the plane and on the sphere simultaneously. The integrability of the Lagrange problem in constant curvature spaces was first discovered by Kozlov and Harin [153], see also [238].

Chapter 6

Contact Geometry and the Restricted Three-Body Problem

6.1 A contact structure for Hill's lunar problem

The following result was proved in [9].

Theorem 6.1.1. *For any given $\mu \in [0,1)$ assume that $c < H(L_1)$, the first critical value of the restricted three-body problem. Then the regularized energy hypersurface $\overline{\Sigma}_c \subset T^*S^2$ of the restricted three-body problem is fiberwise star-shaped.*

For Hill's lunar problem a stronger result was proved by Lee in [158].

Theorem 6.1.2. *Assume that $c < -\frac{3^{\frac{4}{3}}}{2}$. Then the regularized energy hypersurface $\overline{\Sigma}_c \subset T^*S^2$ of Hill's lunar problem is fiberwise convex.*

Remark 6.1.3. It is an open problem if the regularized energy hypersurface of the restricted three-body problem for energies below the first critical value are fiberwise convex.

Remark 6.1.4. Because $\overline{\Sigma}_c$ is fiberwise star-shaped it follows that it is contact. In particular, a result of Cristofaro-Gardiner and Hutchings [65] implies that $\overline{\Sigma}_c$ admits two closed characteristics. We already know the existence of one, namely the retrograde periodic orbit from the work of Birkhoff. Lee's result implies that below the first critical value the regularized energy hypersurfaces of Hill's lunar problem are Finsler and in this case the existence of two closed characteristics was already proved by Bangert and Long [23]. The existence of two closed characteristics in Hill's lunar problem was first proved analytically by Llibre and Roberto in [165].

In this chapter we explain why below the first critical value the energy hypersurface of Hill's lunar problem is fiberwise star-shaped. This is much weaker than the result of Lee [158]. However, the advantage of the proof presented here is

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_6

that the same scheme can also be applied in the restricted three-body problem to prove fiberwise starshapedness [9], although there the argument gets much more involved.

Note that the vector field

$$X = \sum_{i=1}^{2} q_i \frac{\partial}{\partial q_i}$$

is a Liouville vector field on $T^*\mathbb{R}^2$. Indeed, for $\omega = dp \wedge dq$ we obtain by Cartan's formula

$$\mathcal{L}_X \omega = d\iota_X \omega + \iota_X d\omega = d(-qdp) = -dq \wedge dp = dp \wedge dq = \omega.$$

Our next theorem tells us that this Liouville vector field is transverse to Σ_c^b, the bounded component of the energy level $H^{-1}(c)$, and therefore $\iota_X \omega|_{\Sigma_c^b}$ defines a contact structure on Σ_c^b.

Proposition 6.1.5. *Assume that $c < -\frac{3^{\frac{4}{3}}}{2}$. Then $X \pitchfork \Sigma_c^b$, where Σ_c^b is the bounded component of the energy hypersurfaces of Hill's lunar problem to the energy value c, as defined in* (5.31).

Proof. In polar coordinates (r, θ), i.e.,

$$q_1 = r\cos\theta, \quad q_2 = r\sin\theta$$

the Liouville vector field reads

$$X = r\frac{\partial}{\partial r},$$

the effective potential becomes

$$U = -\frac{1}{r} - \frac{3}{2}r^2\cos^2\theta$$

and the Hamiltonian

$$H = \frac{1}{2}\big((p_1 - r\sin\theta)^2 + (p_2 + r\cos\theta)^2\big) + U.$$

To prove the theorem it suffices to show that

$$dH(X)|_{\Sigma_c^b} > 0. \tag{6.1}$$

We estimate using the Cauchy–Schwarz inequality

$$\begin{aligned} dH(X) &= -r\sin\theta(p_1 - r\sin\theta) + r\cos\theta(p_2 + r\cos\theta) + r\frac{\partial U}{\partial r} \\ &\geq r\frac{\partial U}{\partial r} - \sqrt{r^2\sin^2\theta + r^2\cos^2\theta}\sqrt{(p_1 - r\sin\theta)^2 + (p_2 + r\cos\theta)^2} \\ &= r\frac{\partial U}{\partial r} - r\sqrt{2(H-U)}. \end{aligned}$$

This implies that

$$dH(X)|_{\Sigma_c^b} \geq r\left(\frac{\partial U}{\partial r} - \sqrt{2(c-U)}\right).$$

Note that the right-hand side is independent of the variables p_1 and p_2. Therefore to prove (6.1) it suffices to show

$$\left(\frac{\partial U}{\partial r} - \sqrt{2(c-U)}\right)\bigg|_{\mathfrak{K}_c^b} > 0. \tag{6.2}$$

Pick $(r,\theta) \in \mathfrak{K}_c^b$. In particular,

$$U(r,\theta) \leq c.$$

By (8.13) the bounded part of Hill's region is contained in the ball of radius $3^{-\frac{1}{3}}$ centered at the origin. Observe that

$$U|_{\partial B_{3^{-\frac{1}{3}}}(0)} \geq -\frac{3^{\frac{4}{3}}}{2}.$$

Since $c < -\frac{3^{\frac{4}{3}}}{2}$ it follows that there exists

$$\tau \in \left[0, \frac{1}{3^{\frac{1}{3}}} - r\right)$$

such that

$$U(r+\tau,\theta) = c. \tag{6.3}$$

We claim that

$$\frac{\partial U}{\partial r}(q) > 0, \quad q \in B_{3^{-\frac{1}{3}}}(0) \setminus \left\{0, (0,-3^{-\frac{1}{3}}), (0,3^{-\frac{1}{3}})\right\}. \tag{6.4}$$

To see that we estimate

$$\frac{\partial U}{\partial r} = \frac{1}{r^2} - 3r\cos^2\theta \geq \frac{1}{r^2} - 3r \geq 0.$$

If $r < 3^{-\frac{1}{3}}$ the inequality is strict and if $r = 3^{-\frac{1}{3}}$ the first inequality is strict because $\cos^2\theta < 1$ because we removed the points $(0,-3^{-\frac{1}{3}})$ and $(0,3^{-\frac{1}{3}})$ from the ball. This proves (6.4).

We further claim that

$$\frac{\partial^2 U}{\partial r^2}(q) \leq -1, \quad q \in B_{3^{-\frac{1}{3}}}(0) \setminus \{0\}. \tag{6.5}$$

In order to prove that we estimate

$$\frac{\partial^2 U}{\partial r^2} = -\frac{2}{r^3} - 3\cos^2\theta \leq -\frac{2}{r^3} \leq -6 \leq -1.$$

In order to prove (6.2) we estimate using (6.3), (6.4), and (6.5)

$$\begin{aligned}\left(\frac{\partial U}{\partial r}(r,\theta)\right)^2 &= \left(\frac{\partial U}{\partial r}(r+\tau,\theta)\right)^2 - \int_0^\tau \frac{d}{dt}\left(\frac{\partial U(r+t,\theta)}{\partial r}\right)^2 dt \\ &> -2\int_0^\tau \frac{\partial U(r+t,\theta)}{\partial r}\frac{\partial^2 U(r+t,\theta)}{\partial r^2} dt \geq 2\int_0^\tau \frac{\partial U(r+t,\theta)}{\partial r} dt \\ &= 2\big(U(r+\tau,\theta) - U(r,\theta)\big) = 2\big(c - U(r,\theta)\big).\end{aligned}$$

Using (6.4) once more this implies

$$\frac{\partial U}{\partial r}(r,\theta) > \sqrt{2\big(c - U(r,\theta)\big)}.$$

Therefore

$$\left(\frac{\partial U}{\partial r} - \sqrt{2(c-U)}\right)(r,\theta) > 0.$$

This proves (6.2) and the proposition follows. □

Note that the map $(q,p) \mapsto (-p,q)$ is a symplectomorphism of $T^*\mathbb{R}^2$ to itself which interchanges the roles of position and momentum. Hence the base coordinate q becomes the fiber coordinate, whereas the fiber coordinate p becomes the base coordinate after the transformation. Interchanging the roles of q and p in this way the assertion of Proposition 6.1.5 can be interpreted as the fact that Σ_c^b is fiberwise star-shaped in $T^*\mathbb{R}^2$, i.e., if $p \in \mathbb{R}^2$ then the fiber

$$\Sigma_{c,p}^b := T_p^*\mathbb{R}^2 \cap \Sigma_c^b \subset T_p^*\mathbb{R}^2 = \mathbb{R}^2$$

bounds a star-shaped domain.

6.2 Contact connected sum

We briefly recall the connected sum of two smooth manifolds. Suppose that M_1 and M_2 are two oriented n-dimensional manifolds. We will talk about balls D^n which we give their standard orientation. For M_1, choose an embedded ball $\iota_1 : D^n \to M_1$, where ι_1 is orientation preserving. For M_2, choose an embedded ball $\iota_2 : D^n \to M_2$ which reverses orientation. Intuitively, we take out small balls from M_1 and M_2 and glue collar neighborhoods together.

We do this more precisely. Write D_r for the ball with radius r, so

$$D_r = \{z \in D^n \mid \|z\| < r\}.$$

Fix a number R with $0 < R < 1$. We will use the annulus $A := D^n - \overline{D_R}$, and the orientation reversing map

$$\begin{aligned}\rho : A &\longrightarrow A \\ x &\longmapsto (1 + R - \|x\|) \cdot \frac{x}{\|x\|}.\end{aligned}$$

In addition, the map reverses the inner and outer sphere of the annulus, meaning the spheres with radii R and 1, respectively. We define the *connected sum* of M_1 and M_2 as

$$M_1 \# M_2 := M_1 \setminus \iota_1(\overline{D_R}) \coprod M_2 \setminus \iota_2(\overline{D_R}) / \sim$$

where $\sim$ is an equivalence relation. Namely, if $\tilde{x} \in M_1$ is given by $\tilde{x} = \iota_1(x)$, and $\tilde{y} \in M_2$ is given by $\tilde{y} = \iota_2(y)$, then we say that $\tilde{x} \sim \tilde{y}$ if and only if $\rho(x) = y$. Other points are not related. Geometrically, the above just means that we glue the two annuli $\iota_1(A)$ and $\iota_2(A)$ together by reversing inner and outer spheres.

Lemma 6.2.1. *Suppose that M_1 and M_2 are oriented smooth n-dimensional manifolds. Then the connected sum $M_1 \# M_2$ as defined above is an oriented, smooth manifold.*

A detailed proof can be found in Chapter 10 of [44]. Here we give an outline of the main ingredients.

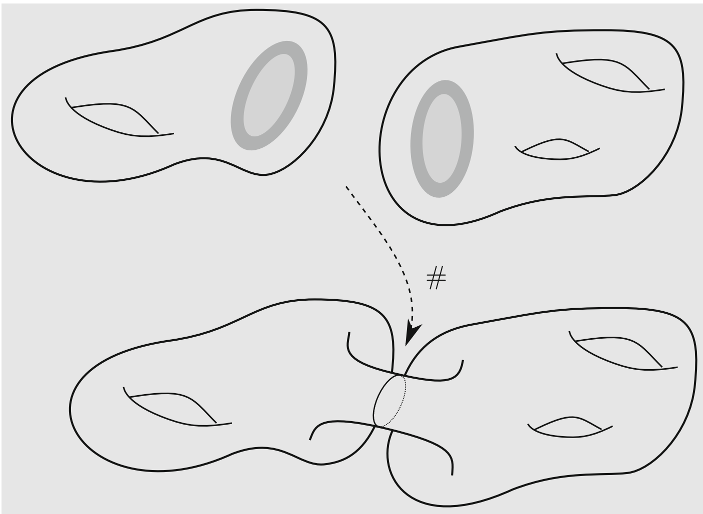


Figure 6.1: Connected sum of smooth, oriented manifolds.

Outline of the proof. Clearly $M_1 \setminus \iota_1(\overline{D_R})$ and $M_2 \setminus \iota_2(\overline{D_R})$ are smooth manifolds, and the above equivalence relation glues diffeomorphically along open sets, so we obtain a topological space which is locally Euclidean. Furthermore, the choice of the map ρ also ensures that the quotient space is Hausdorff, so we obtain the structure of a manifold on $M_1 \# M_2$. A smooth atlas can be constructed directly from the smooth atlases on both M_1 and M_2: the only new case concerns transition functions for charts with non-empty intersection with both M_1 and M_2. These transition functions are also smooth due to the gluing via a diffeomorphism. To see that we get an orientation, we observe that ι_2 and ρ are orientation reversing, so their composition is orientation preserving. □

Remark 6.2.2. The connected sum is actually well defined, but this requires an additional argument involving the so-called disk theorem, which asserts that two embeddings of closed n-disks in a connected n-manifold are ambiently isotopic, provided the induced orientations on the disk coincide; using this theorem one can define a (trivial) cobordism between the connected sum of M_1 and M_2 with one set of choices of embedded disks and another, showing that the connected sum is well defined.

6.2.1 Contact version

The contact version mimics the above construction, but instead of gluing the two annuli directly together, we define a model for the connecting tube. The discussion here follows Weinstein's ideas, and we have drawn a picture describing the construction of the tube in Figure 6.2. This model will appear again when we

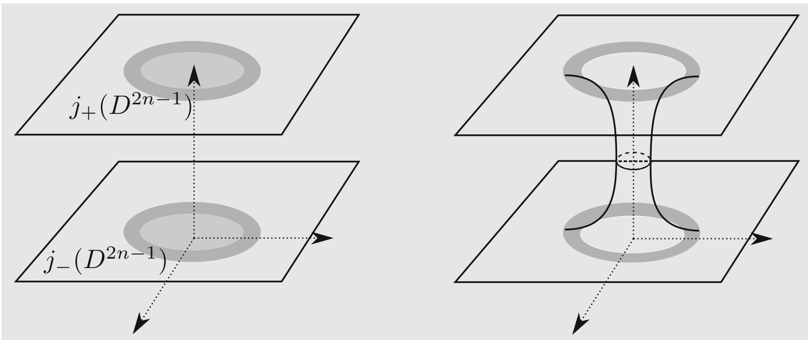


Figure 6.2: The tube of the contact connected sum.

look at level sets of Hamiltonians.

Consider contact manifolds (M_1^{2n-1}, α_1) and (M_2^{2n-1}, α_2) together with two points, say $q_1 \in M_1$ and $q_2 \in M_2$. We want to define the contact connected sum. As before, choose a number $0 < R < 1$ for the radius of the ball we are going to cut out.

Choose Darboux balls $\iota_k : D^{2n-1} \to M_k$ containing q_k for $k = 1, 2$; these are provided by Theorem 2.5.3. To construct the tube, we also embed these Darboux balls into $\mathbb{R}^{2n}$ by the map

$$\begin{aligned} j_\pm : D^{2n-1} &\longrightarrow \mathbb{R}^{2n} \\ (x, y; z) &\longmapsto (x, y, z, \pm 1). \end{aligned}$$

We observe that the vector field

$$X = \frac{1}{2}(x \cdot \partial_x + y \cdot \partial_y) + 2z\partial_z - w\partial_w$$

is a Liouville vector field and that it is transverse to $j_\pm(D^{2n-1})$. Furthermore, it induces the Liouville form

$$i_X\omega_0 = \frac{1}{2}(xdy - ydx) + 2zdw + wdz,$$

which restricts to the contact form $\frac{1}{2}(xdy - ydx) + dz$ on $j_+(D^{2n-1})$ and to the contact form $\frac{1}{2}(xdy - ydx) - dz$ on $j_-(D^{2n-1})$. These are the standard contact form with the standard orientation, and a variation of the standard contact form with the *opposite* orientation. To construct the tube, for $0 < \epsilon < R < 1$ choose a smooth, increasing function $f : \mathbb{R} \to \mathbb{R}$ with the following properties

- $f(z) = 1$ if $z > R$
- $f(0) < 0$ and $f(\epsilon) = 0$ and $f'(\epsilon) > 0$.

We define the connecting tube as the set

$$\mathcal{T} := \{(x, y; z, w) \in \mathbb{R}^{2n-2} \times \mathbb{R}^2 \mid w^2 = f(|x|^2 + |y|^2 + z^2)\} \cap \{|x|^2 + |y|^2 + z^2 < 1\}.$$

Lemma 6.2.3. *The connecting tube is a smooth submanifold of* $\mathbb{R}^{2n}$. *Furthermore, the Liouville vector field* $X = \frac{1}{2}(x\partial_x + y\partial_y) + 2z\partial_z - w\partial_w$ *is transverse to* $\mathcal{T}$, *so* $\mathcal{T}$ *is a contact manifold. For* $(x, y; z)$ *with* $|x|^2 + |y|^2 + z^2 > R$, *the contact form coincides with the standard contact form on* $j_\pm(D^{2n-1})$.

Proof. Since $\mathcal{T}$ is a level set of the function $F : (x, y; z, w) \mapsto f(|x|^2+|y|^2+z^2)-w^2$, it suffices to check that $X(F) \neq 0$. Indeed, this shows that dF is never zero, and of course it shows that X is transverse to $\mathcal{T}$. We now compute

$$X(F) = f'(|x|^2 + |y|^2 + z^2) \cdot (|x|^2 + |y|^2 + 4z^2) + 2w^2.$$

Since f is increasing, all terms are non-negative. To see that their sum is always positive on $\mathcal{T}$, we note the following.

- If $w \neq 0$, then $2w^2$ is positive.
- If $w = 0$, then $f(|x|^2 + |y|^2 + z^2) = 0$ and $|x|^2 + |y|^2 + z^2 = \epsilon$, and by assumption $f'(\epsilon) > 0$.

The last claim follows since $\mathcal{T}$ and $\iota_j(D^{2n-1})$ coincide if $|x|^2 + |y|^2 + z^2 > R$, and their contact forms are induced by the same Liouville vector field. □

We now define the contact connected sum by

$$(M_1, \alpha_1)\#(M_2, \alpha_2) := M_1 \setminus \iota_1(\overline{D_R}) \coprod M_2 \setminus \iota_2(\overline{D_R}) \coprod \mathcal{T}/\sim.$$

Here the equivalence relation is defined as follows.

- If $\tilde{x} = \iota_1(x)$ lies in $M_1 \setminus \iota_1(\overline{D_R})$ and $\tilde{y} = j_1(y)$ lies in $\mathcal{T}$, then $\tilde{x} \sim \tilde{y}$ if and only if $x = y$.
- If $\tilde{x} = \iota_2(x)$ lies in $M_2 \setminus \iota_2(\overline{D_R})$ and $\tilde{y} = j_2(y)$ lies in $\mathcal{T}$, then $\tilde{x} \sim \tilde{y}$ if and only if $x = y$.
- Other pairs of points are not related.

Theorem 6.2.4. *The space* $(M_1, \alpha_1)\#(M_2, \alpha_2)$ *is a contact manifold.*

Proof. The proof that this is a smooth manifold follows the same argument as before: since we glue diffeomorphically along open sets, we clearly get a differentiable atlas. Inspecting the equivalence relation we also see that the Hausdorff property holds. To see that this is a contact manifold, we only need to observe that $M_1 \setminus \iota_1(\overline{D_R})$, $M_2 \setminus \iota_2(\overline{D_R})$ and $\mathcal{T}$ are contact manifolds with contact forms that patch together with the gluing relation. This was checked in the above. □

Remark 6.2.5. The contact connected sum is also well defined. The main ingredient here is the contact disk theorem, see Theorem 2.6.7 in [97].

6.3 A real contact structure for the restricted three-body problem

Recall that the Hamiltonian of the restricted three-body problem

$$H(q,p) = \frac{1}{2}|p|^2 - \frac{\mu}{|q-m|} - \frac{1-\mu}{|q-e|} + q_1 p_2 - q_2 p_1$$

is invariant under the anti-symplectic involution $\rho\colon T^*\mathbb{R}^2 \to T^*\mathbb{R}^2$ given by

$$\rho_0(q_1, q_2, p_1, p_2) = (q_1, -q_2, -p_1, p_2).$$

As we discussed in Theorem 6.1.1 for every $\mu \in [0,1)$ and every $c < H(L_1)$, the first critical value, the regularized energy hypersurface

$$\overline{\Sigma}_c \subset T^*S^2$$

is fiberwise starshaped so that the restriction of the Liouville one-form on T^*S^2 gives a contact form on $\overline{\Sigma}_c$. Let

$$\rho\colon T^*S^2 \to T^*S^2$$

be the anti-symplectic involution on T^*S^2 obtained in the following way. On $S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$ consider the great circle $\{x \in S^2 : x_1 = 0\}$ and let $I\colon S^2 \to S^2$ be the reflection at this circle. Then the symplectic involution $d_*I\colon T^*S^2 \to T^*S^2$ commutes with the anti-symplectic involution $(p,q) \mapsto (p,-q)$ on T^*S^2, so that their composition is again an anti-symplectic involution which we denote by ρ. The fact that the Hamiltonian is invariant under ρ_0 translates into the fact that $\overline{\Sigma}_c$ is invariant under ρ. This leads to interesting relations between the restricted three-body problem and Dihedral homology [92, 166].

Chapter 7

Periodic Orbits in Hamiltonian Systems

7.1 A short history of the research on periodic orbits

Very early pioneers in the search of periodic orbits were Euler and Lagrange. They both found *homographic* solutions to the planar three-body problem. This means a solution whose position coordinates remain self-similar to its initial condition as time varies. More precisely, if $q_i(t)$ denotes the position of the ith particle at time t with center of mass at the origin, then

$$q_i(t) = r(t) \cdot R(\theta(t)\,)q_i(0)$$

for a time-dependent rotation $R(\theta(t)\,)$. Euler found collinear homographic solutions to the planar three-body problem and Lagrange found other types of homographic solutions, namely initial configurations where the points $q_i(0)$ form the vertices of an equilateral triangle. These types of special initial conditions that lead to self-similar solutions are known as central configurations. This is a large and interesting topic in celestial mechanics which we will not touch. However, see for example [181] for lecture notes and [113] for an interesting recent study of these topics. To get an idea how these homographic orbits of Euler and Lagrange look like, we refer to Figure 7.1. We shall see special cases of these solutions in the restricted three-body problem in the form of the five Lagrange points.

The study of special periodic solutions in the three-body problem or, more generally, in the N-body problem has actually seen huge developments in past two decades due to the discovery of *choreographies*, starting with Moore's discovery in [182] through numerical means and their rediscovery by Chenciner and Montgomery in [52] through both analytic and numerical means. See also Simó's work, [223] for a great variety of choreographies. Simply put, choreographies are solutions to the N-body problem where the N bodies follow each other on a given closed curve. The simplest choreographies are special cases of Lagrange's equilateral solutions, so in this sense choreographies generalize some homographic solutions.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_7

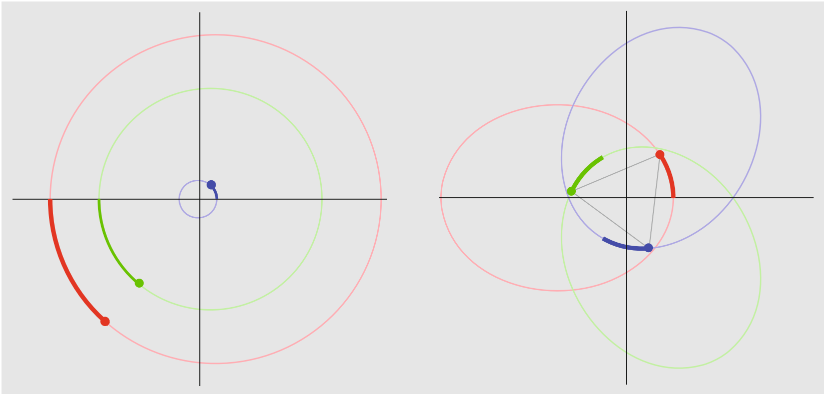


Figure 7.1: Homographic solutions coming from collinear and equilateral central configurations.

The pioneer in the search for periodic orbits in the *restricted* three-body problemwas Hill with his discovery of the retrograde and direct periodic orbit in Hill's lunar problem in [116, 117]. His motivation was to describe the motion of the moon. Therefore the big mass in Hill's lunar theory is interpreted as the sun, the intermediate mass as the earth and the satellite is the moon. The physical moon moves in a direct or prograde manner around the earth, i.e., it moves in the same direction as the earth rotates around the sun. Therefore the direct periodic orbit found by Hill is a good approximation of the actual movement of the moon. Of course Hill's lunar problem is a simplification of the actual situation. For example it ignores the eccentricity of the orbit of the earth and because after blowing up the coordinates the sun is infinitely far away its parallax vanishes. Moreover, the mass distribution of the earth and the influence of the other planets on the moon are of course not considered in Hill's lunar problem. However, Hill's idea was to use the direct periodic orbit, as an *intermediate orbit* of the actual orbit of the physical moon. Namely to obtain the orbit of the actual physical moon one thinks of the ignored effects as perturbations and then applies perturbation theory to the direct periodic orbit to obtain better and better approximations of the actual orbit of the physical moon. This idea turned out to be quite successful and the convergence of the perturbation series obtained from Hill's procedure turned out to be much faster than the convergence of other perturbative approaches to the orbit of the moon like the sophisticated theory of Delaunay [70, 71, 72] which, different from Hill's theory, started from a Kepler ellipse as intermediate orbit for the actual orbit of the moon. More information on the different approaches to the Lunar theory can be found in the classical book by Brown [48]. For a more recent treatment of this topic, see for example [192, Chapter 22]. A discussion of Hill's approach to the retrograde and direct periodic orbit in Hill's lunar problem can also be found in the book by Wintner [243, Chapter VI]. Hill found actually a trigonometric series for the orbits. The question of convergence of this trigonometric series is discussed as well in Wintner's book. The anxious reader can be relieved. For the values of our physical moon Hill's trigonometric series of the direct periodic orbit converges as Wintner has shown in his thesis [242].

Poincaré was extremely impressed by Hill's work to such an extent that according to Ginsburg and Smith [99, page 131], when he met Hill, he greeted him with the words "You are the one man I came to America to see". It was Poincaré who stressed the importance of periodic orbits. Let us listen to his own words in Section 36 of Chapter III of the first volume of his famous "Méthodes Nouvelles de la Mécanique Céleste" [205], in the translation of Chenciner [54]

> Indeed, there is zero probability that the initial conditions of motion be precisely those which correspond to a periodic solution. But it may happen that they differ very little, and this occurs precisely where the old methods no longer apply. We can then use the periodic solution as a first approximation, as an *intermediate orbit*, to use the language of Mr. Gyldén. There is even more: here is a fact that I could not prove rigorously, but which nevertheless seems very likely to me. Given equations of the form defined in Section 13 and an arbitrary solution of these equations, one can always find a periodic solution (with a period which, admittedly, may be very long), such that the difference between the two solutions be arbitrarily small. In fact, what makes these solutions so precious to us, is that they are, so to say, the only opening through which we can try to enter a place which, up to now, was deemed inaccessible.

If one is generous with the topology and if one is allowed to perturb the Hamiltonian slightly, Poincaré's expectation about denseness of periodic orbits turned out to be true as was shown by Pugh and Robinson [208]. Apart from that Poincaré's quest for periodic orbits turned out to be a very fruitful stimulus of mathematical research. Already in the simplest case, namely the free particle where the Hamiltonian just consists of kinetic energy, this question turned out to be highly interesting. As we have seen in Section 2.3.1 of Chapter 2 the search for periodic orbits for the free particle corresponds to the search for closed geodesics. Already Poincaré pioneered the search for closed geodesics with his theorem about existence of a closed geodesic on convex surfaces [206]. Interesting information about the historical context as well as the philosophy and epistemology of Poincaré can be found in [27, 210].

By interpreting closed geodesics as critical points of the energy functional on the free loop space Morse developed Morse theory [185] in order to detect closed geodesics. In particular, this showed that there are interesting relations between the question on the existence of closed geodesics and the topology of the free loop space. The approach of Morse also works for Finsler metrics. It was Floer with his discovery of a semi-infinite-dimensional Morse theory [85], nowadays referred to as Floer homology, who taught us how methods from Morse theory can be used to find periodic orbits for general Hamiltonian systems.

7.2 Variational approach

Assume that (M, ω) is a symplectic manifold and $H \in C^\infty(M, \mathbb{R})$ is a Hamiltonian with the property that 0 is a regular value, i.e., $\Sigma = H^{-1}(0) \subset M$ is a regular hypersurface. Abbreviate by $S^1 = \mathbb{R}/\mathbb{Z}$ the circle. A *parametrized periodic orbit* on Σ is a loop $\gamma \in C^\infty(S^1, \Sigma)$ for which there exists $\tau > 0$ such that the pair (γ, τ) is a solution of the problem

$$\partial_t \gamma(t) = \tau X_H(\gamma(t)), \quad t \in S^1.$$

Because γ is parametrized the positive number $\tau = \tau(\gamma)$ is uniquely determined by γ. We refer to τ as the *period* of γ. Indeed, if we reparametrize γ to $\gamma_\tau \colon \mathbb{R} \to \Sigma$ by

$$\gamma_\tau(t) = \gamma\big(\tfrac{t}{\tau}\big)$$

then γ_τ satisfies

$$\partial_t \gamma_\tau(t) = X_H(\gamma_\tau(t)), \quad \gamma_\tau(t+\tau) = \gamma(t),\ t \in \mathbb{R}.$$

We next explain how parametrized periodic orbits can be interpreted variationally as critical points of an action functional. To simplify the discussion we assume that (M, ω) is an exact symplectic manifold, i.e., ω admits a primitive λ such that $d\lambda = \omega$. Abbreviate by

$$\mathfrak{L} = C^\infty(S^1, M)$$

the free loop space of M and let $\mathbb{R}_+ = (0, \infty)$ denote the positive real numbers. Consider

$$\mathcal{A}^H \colon \mathfrak{L} \times \mathbb{R}_+ \to \mathbb{R},$$

defined for a free loop $\gamma \in \mathfrak{L}$ and $\tau \in \mathbb{R}_+$ by

$$\mathcal{A}^H(\gamma, \tau) = \int_{S^1} \gamma^* \lambda - \tau \int_{S^1} H(\gamma(t))dt. \tag{7.1}$$

One might think of $\mathcal{A}^H$ as the Lagrange multiplier functional of the area functional of the constraint given by the mean value of H. We refer to $\mathcal{A}^H$ as *Rabinowitz action functional*.

Remark 7.2.1. The action functional $\mathcal{A}^H$ has itself an interesting history. Although Moser explicitly wrote in 1976, [187], that $\mathcal{A}^H$ is useless in finding periodic orbits, it was used shortly after in 1978 by Rabinowitz in his celebrated paper [209] to prove existence of periodic orbits on star-shaped hypersurfaces in $\mathbb{R}^{2n}$ and so opened the way for the application of global methods in Hamiltonian mechanics. Moreover, the action functional $\mathcal{A}^H$ can be used to define a semi-infinite-dimensional Morse homology in the sense of Floer [85], which is referred to as Rabinowitz Floer homology, see [8, 55].

Lemma 7.2.2. *Critical points of $\mathcal{A}^H$ precisely consist of pairs (γ, τ), where γ is a parametrized periodic orbit of X_H on $\Sigma = H^{-1}(0)$ of period τ.*

Proof. Given a loop $\gamma \in \mathfrak{L}$, the tangent space of $\mathfrak{L}$ at γ

$$T_\gamma \mathfrak{L} = \Gamma(\gamma^* TM)$$

consists of vector fields along γ. Suppose that $(\gamma, \tau) \in \mathfrak{L} \times \mathbb{R}_+$ and pick

$$(\widehat{\gamma}, \widehat{\tau}) \in T_{(\gamma,\tau)}(\mathfrak{L} \times \mathbb{R}_+) = T_\gamma \mathfrak{L} \times \mathbb{R}.$$

By applying Cartan's formula to the Lie derivative $\mathcal{L}_{\widehat{\gamma}} \lambda$ we compute for the pairing of the differential of $\mathcal{A}^H$ with $(\widehat{\gamma}, \widehat{\tau})$

$$\begin{aligned}
d\mathcal{A}^H_{(\gamma,\tau)}(\widehat{\gamma}, \widehat{\tau}) &= \int_{S^1} \gamma^* \mathcal{L}_{\widehat{\gamma}} \lambda - \tau \int_{S^1} dH(\gamma)\widehat{\gamma} dt - \widehat{\tau} \int_{S^1} H(\gamma) dt \\
&= \int_{S^1} \gamma^* d\iota_{\widehat{\gamma}} \lambda + \int_{S^1} \gamma^* \iota_{\widehat{\gamma}} d\lambda - \tau \int \omega(\widehat{\gamma}, X_H(\gamma)) dt \\
&\quad - \widehat{\tau} \int_{S^1} H(\gamma) dt \\
&= \int_{S^1} d\gamma^* \iota_{\widehat{\gamma}} \lambda + \int_{S^1} \gamma^* \iota_{\widehat{\gamma}} \omega - \int \omega(\widehat{\gamma}, \tau X_H(\gamma)) dt \\
&\quad - \widehat{\tau} \int_{S^1} H(\gamma) dt \\
&= \int_{S^1} \omega(\widehat{\gamma}, \partial_t \gamma) dt - \int \omega(\widehat{\gamma}, \tau X_H(\gamma)) dt - \widehat{\tau} \int_{S^1} H(\gamma) dt \\
&= \int_{S^1} \omega(\widehat{\gamma}, \partial_t \gamma - \tau X_H(\gamma)) dt - \widehat{\tau} \int_{S^1} H(\gamma) dt.
\end{aligned}$$

We conclude that a critical point (γ, τ) of $\mathcal{A}^H$ is a solution of the problem

$$\begin{cases} \partial_t \gamma - \tau X_H(\gamma) = 0 \\ \int_{S^1} H(\gamma) dt = 0. \end{cases} \tag{7.2}$$

In view of preservation of energy, as explained in Theorem 2.2.2, problem (7.2) is equivalent to the following problem

$$\begin{cases} \partial_t \gamma = \tau X_H(\gamma) \\ H(\gamma) = 0, \end{cases} \tag{7.3}$$

i.e., the mean value constraint can be replaced by a pointwise constraint. But solutions of problem (7.3) are precisely parametrized periodic orbits of X_H of period τ on the energy hypersurface $\Sigma = H^{-1}(0)$. This proves the lemma. $\square$

There are two actions on the free loop space $\mathfrak{L} = C^\infty(S^1, M)$. The first action comes from the group structure of the domain S^1. Given $r \in S^1$ and $\gamma \in \mathfrak{L}$ we reparametrize γ by

$$r_*\gamma(t) = \gamma(r+t), \quad t \in S^1.$$

If we extend this S^1-action of $\mathfrak{L}$ to $\mathfrak{L} \times \mathbb{R}$ by acting trivially on the second factor, the action functional $\mathcal{A}^H$ is S^1-invariant. Its critical set, namely the set of parametrized periodic orbits, is then S^1-invariant as well, which is also obvious from the ODE that parametrized periodic orbits satisfy. We refer to an equivalence class of parametrized periodic orbits under reparametrization by S^1 as an *unparametrized periodic orbit*.

The second action on the free loop space comes from the fact that S^1 is diffeomorphic to its finite covers. Consider the action of the monoid $\mathbb{N}$ on $\mathfrak{L}$ given for $n \in \mathbb{N}$ and $\gamma \in \mathfrak{L}$ by

$$n_*\gamma(t) = \gamma(nt), \quad t \in S^1.$$

The two actions combined give rise to an action on $\mathfrak{L}$ of the semi-direct product $\mathbb{N} \ltimes S^1$ with product defined as

$$(n_1, r_1)(n_2, r_2) = (n_1 n_2, n_1 r_2 + r_1).$$

We extend the action of $\mathbb{N}$ on $\mathfrak{L}$ to an action of $\mathbb{N}$ on $\mathfrak{L} \times \mathbb{R}$ which is given for $\gamma \in \mathfrak{L}$, $\tau \in \mathbb{R}$ and $n \in \mathbb{N}$ by

$$n_*(\gamma, \tau) = (n_*\gamma, n\tau).$$

The action functional $\mathcal{A}^H$ is homogeneous of degree one for the action by $\mathbb{N}$, i.e.,

$$\mathcal{A}^H(n_*(\gamma, \tau)) = n\mathcal{A}^H(\gamma, \tau), \qquad (\gamma, \tau) \in \mathfrak{L} \times \mathbb{R}, \quad n \in \mathbb{N}.$$

Therefore its critical set is invariant under the action of $\mathbb{N}$, too, again a fact which can immediately be understood as well by looking at the ODE. A parametrized periodic orbit γ is called *multiply covered* if there exists a parametrized periodic orbit γ_1 and a positive integer $n \geq 2$ such that $\gamma = n_*\gamma_1$. Note that this notion does not depend on the parametrization of the orbit such that one can talk about multiple covers on the level of unparametrized periodic orbits. A parametrized or unparametrized periodic orbit is called *simple* if it is not multiply covered. Because a parametrized periodic orbit is a solution to a first-order ODE, it follows that for a simple periodic orbit it holds that

$$\gamma(t) = \gamma(t') \quad \Longleftrightarrow \quad t = t' \in S^1.$$

Moreover, for every parametrized periodic orbit γ there exists a unique simple periodic orbit γ_1 and a unique $k \in \mathbb{N}$ such that $\gamma = k_*\gamma_1$. We refer to k as the *covering number* of the periodic orbit γ. Alternatively, the covering number can also be defined as

$$\operatorname{cov}(\gamma) = \max\Big\{k \in \mathbb{N} : \gamma(t + \tfrac{1}{k}) = \gamma(t), \ \forall\, t \in S^1\Big\}.$$

Again the covering number does not depend on the parametrization and can therefore be associated to unparametrized periodic orbits as well. With the notion of the covering number at our disposal we can characterize simple periodic orbits as the periodic orbits of covering number one. The period of the underlying simple periodic orbit of an (un)parametrized periodic orbit γ is referred to as the *minimal period* which we denote by $\tau_0(\gamma)$. Note that we have the relation

$$\tau(\gamma) = \mathrm{cov}(\gamma) \cdot \tau_0(\gamma).$$

By a *periodic orbit* we mean the trace $\{\gamma(t) : t \in S^1\}$ of a parametrized periodic orbit. Note that there is a one-to-one correspondence between periodic orbits and simple unparametrized orbits, or put differently, equivalence classes of parametrized periodic orbits under the action of the monoid $\mathbb{N} \ltimes S^1$. The *period* of the periodic orbit is then the period of the underlying simple periodic orbit, respectively the minimal period of all of its representatives. If we think of $(\Sigma, \omega|_\Sigma)$ as a Hamiltonian manifold, a periodic orbit corresponds to a closed leaf of the foliation $\ker \omega$ on Σ. However, note that the period of the periodic orbit only makes sense with reference to the Hamiltonian H and cannot be determined directly from the Hamiltonian structure $\omega|_\Sigma$.

7.3 The kernel of the Hessian

Because parametrized periodic orbits can always be reparametrized, they never occur as isolated points in the free loop space. In terms of the action functional this means that $\mathcal{A}^H$ is never Morse, i.e., the kernel of the Hessian of $\mathcal{A}^H$ at a critical point is never just the zero vector space. This is a common phenomenon for functionals which are invariant under a continuous symmetry because the infinitesimal generators of the symmetry belong to the kernel of the Hessian.

Assume that $(\widehat{\gamma}, \widehat{\tau})$ lies in the kernel of the Hessian of $\mathcal{A}^H$ at the critical point (γ, τ) of $\mathcal{A}^H$. We define the path $v\colon [0,1] \to T_{\gamma(0)}M$ by

$$v(t) = d\phi_H^{-\tau t}(\gamma(t))\widehat{\gamma}(t). \tag{7.4}$$

Because $\widehat{\gamma}$ is a loop, we get

$$v(1) = d\phi_H^{-\tau}(\gamma(1))\widehat{\gamma}(1) = d\phi_H^{-\tau}(\gamma(0))\widehat{\gamma}(0) = d\phi_H^{-\tau}(\gamma(0))v(0)$$

so that v satisfies the twist condition

$$v(0) = d\phi_H^{\tau}(\gamma(0))v(1). \tag{7.5}$$

The assumption that $(\widehat{\gamma}, \widehat{\tau})$ lies in the kernel of the Hessian of $\mathcal{A}^H$ at (γ, τ) is equivalent to the problem

$$\begin{cases} \partial_t v(t) = \widehat{\tau} X_H(\gamma(0)), \quad t \in [0,1], \\ \int_0^1 dH(\gamma(0))v(t)dt = 0. \end{cases} \tag{7.6}$$

Integrating the first equation we obtain

$$v(t) = v(0) + t\widehat{\tau} X_H(\gamma(0)).$$

Plugging this into the second equation we obtain using that $dH(X_H) = 0$

$$\int_0^1 dH(\gamma(0))v(0)dt = 0$$

but now the integrand does not depend on t anymore, so that we can rewrite this as

$$dH(\gamma(0))v(0) = 0.$$

In view of (7.5) we see that by abbreviating $v_0 = v(0)$ problem (7.6) is equivalent to the problem

$$\begin{cases} d\phi_H^{-\tau}(\gamma(0))v_0 = v_0 + \widehat{\tau} X_H(\gamma(0)), \\ \quad dH(\gamma(0))v_0 = 0. \end{cases} \tag{7.7}$$

If γ is a parametrized periodic orbit on the energy hypersurface $H^{-1}(0)$ of period τ, we abbreviate

$$\mathfrak{K}(\gamma) := \{(v_0, \widehat{\tau}) \in T_{\gamma(0)}M \times \mathbb{R} : \text{solution of } (7.7)\}.$$

The discussion above shows that the vector space $\mathfrak{K}(\gamma)$ is canonically isomorphic to the kernel of the Hessian of $\mathcal{A}^H$ at (γ, τ). Because the action functional $\mathcal{A}^H$ is invariant under the S^1-action obtained by rotating the loop, the kernels of the Hessians at two critical points belonging to the same S^1-orbit are canonically isomorphic. If $r \in S^1$, then the isomorphism

$$\Psi_r \colon \mathfrak{K}(\gamma) \to \mathfrak{K}(r_*\gamma) \tag{7.8}$$

can be explicitly written down, namely

$$\Psi_r(v_0, \widehat{\tau}) = \big(d\phi_H^{r\tau}(\gamma(0))v_0 + r\widehat{\tau} X_H(\gamma(r\tau)), \widehat{\tau}\big).$$

Note that strictly speaking the two summands in the first component are only well defined by choosing a representative of $r \in S^1 = \mathbb{R}/\mathbb{Z}$ in the covering space $\mathbb{R}$. However, we claim the sum is independent of this choice of representative. To see this, we compute with the first equation in (7.7) and $\phi_H^\tau(\gamma(0)) = \gamma(\tau) = \gamma(0)$ in mind that

$$\begin{aligned} X_H(\gamma(0)) &= X_H(\gamma(\tau)) \\ &= \frac{d}{dt}\Big|_{t=0} \phi_H^{\tau+t}(\gamma(0)) \\ &= \frac{d}{dt}\Big|_{t=0} \phi_H^\tau \phi_H^t(\gamma(0)) \\ &= d\phi^\tau(\gamma(0))X_H(\gamma(0)). \end{aligned} \tag{7.9}$$

This equality shows also that always

$$\big(X_H(\gamma(0)), 0\big) \in \mathfrak{K}(\gamma).$$

This is no surprise at all, but follows again from the fact that the action functional $\mathcal{A}^H$ is invariant under the S^1 action on $\mathfrak{L} \times (0, \infty)$ obtained by rotating the loop and the solution $(X_H(\gamma(0)), 0)$ of (7.7) is the infinitesimal generator of this action.

Definition 7.3.1. A parametrized periodic orbit γ is called *(transversely) non-degenerate* if

$$\mathfrak{K}(\gamma) = \langle (X_H(\gamma(0)), 0) \rangle.$$

In view of the isomorphism (7.8) the notion of non-degeneracy does not depend on the choice of parametrization and we can talk as well of non-degeneracy on the level of unparametrized periodic orbits. The non-degeneracy condition of a periodic orbit might be interpreted as a Morse–Bott condition of the action functional $\mathcal{A}^H$. In general a smooth function is called *Morse–Bott* if its critical set is a manifold and the tangent space to this manifold is given by the kernel of the Hessian. In the case at hand the non-degenerate unparametrized periodic orbit, namely the S^1-family of parametrized periodic orbits obtained by reparametrization, is a manifold, more precisely a circle, and its tangent space coincides with the kernel of the Hessian. Although reparametrizations of a non-degenerate periodic orbit are non-degenerate as well, multiple covers of a non-degenerate parametrized periodic orbit do not need to be non-degenerate. In fact, if γ is a parametrized periodic orbit and $n \in \mathbb{N}$, then we have an injection

$$\Psi_n \colon \mathfrak{K}(\gamma) \to \mathfrak{K}(n_* \gamma)$$

which does not need to be a surjection. This injection is given by the formula

$$\Psi_n(v_0, \widehat{\tau}) = (v_0, n\widehat{\tau}).$$

In view of the injection Ψ_n we can say that the underlying simple orbit of a non-degenerate parametrized or unparametrized orbit is again non-degenerate but the converse does not follow. We therefore make the following definition

Definition 7.3.2. A periodic orbit is *non-degenerate*, if each of its parametrized representatives is non-degenerate in the sense of Definition 7.3.1.

We can characterize the non-degeneracy of a parametrized periodic orbit a bit differently by introducing the following notation. We abbreviate by $\Sigma = H^{-1}(0)$ the energy hypersurface and recall that the pair (Σ, ω) is a Hamiltonian manifold whose one-dimensional distribution $\ker(\omega)$ is spanned by $X_H|_\Sigma$. Because $d\phi_H^\tau$ leaves $\ker(\omega)$ invariant by (7.9) we get an induced bundle map $\overline{d\phi_H^\tau} \colon T\Sigma/\ker(\omega) \to T\Sigma/\ker(\omega)$ which is characterized by the following commutative diagram

$$\begin{tikzcd} T\Sigma \arrow[r, "d\phi_H^\tau"] \arrow[d, "\pi"] & T\Sigma \arrow[d, "\pi"] \\ T\Sigma/\ker(\omega) \arrow[r, "\overline{d\phi_H^\tau}"] & T\Sigma/\ker(\omega) \end{tikzcd} \tag{7.10}$$

where $\pi\colon T\Sigma \to T\Sigma/\ker(\omega)$ is the projection along $\ker(\omega)$. Note that $T\Sigma/\ker(\omega)$ is a symplectic vector bundle over Σ and $\overline{d\phi_H^\tau}$ is a symplectic bundle map. Then a parametrized periodic orbit γ is non-degenerate if and only if

$$\ker\left(\overline{d\phi_H^\tau}(\gamma(0)) - \mathrm{id}|_{T_{\gamma(0)}\Sigma/\ker(\omega)}\right) = \{0\}, \tag{7.11}$$

i.e., $\overline{d\phi_H^\tau}(\gamma(0))$ has no eigenvalue equal to one.

If the energy hypersurface $\Sigma = H^{-1}(0)$ admits a contact form λ such that

$$X_H|_\Sigma = R,$$

where R is the Reeb vector field of (Σ, λ), then the discussion about non-degeneracy simplifies. To characterize the kernel of the Hessian we decompose

$$v(t) = v_\xi(t) + a(t)R(\gamma(0))$$

where

$$v_\xi(t) \in \xi_{\gamma(t)}, \quad a(t) \in \mathbb{R},\ t \in [0,1]$$

for $\xi = \ker\lambda$ the hyperplane distribution. Because

$$d\phi_R^\tau(\gamma(0)) = (d^\xi\phi_R^\tau(\gamma(0)), \mathrm{id})\colon \xi_{\gamma(0)} \oplus \langle R_{\gamma(0)}\rangle \to \xi_{\gamma(0)} \oplus \langle R_{\gamma(0)}\rangle$$

we conclude that

$$\widehat{\tau} = 0, \quad d^\xi\phi^\tau(\gamma(0))v_\xi(0) = v_\xi(0).$$

Therefore in the contact case we can identify the kernel of the Hessian of $\mathcal{A}^H$ at (γ, τ) with $\ker\left(d^\xi\phi^\tau(\gamma(0)) - \mathrm{id}|_{\xi_{\gamma(0)}}\right) \oplus \langle R_{\gamma(0)}\rangle$.

One can show that, given any Hamiltonian function, there is a perturbation, which can be made arbitrarily small in the C^∞-topology such that the periodic orbits of the perturbed Hamiltonian are all non-degenerate in the above sense for a fixed energy level. However, there are many important examples of Hamiltonian systems where there are degenerate periodic orbits. Often this happens because one studies a Hamiltonian which is invariant under some symmetry. Then the symmetry also acts on periodic orbits and therefore periodic orbits are not isolated anymore. On the other hand after a small perturbation the Hamiltonian cannot be expected to be still invariant under the symmetry. Alas, there are only a few examples where we completely understand the dynamics and these examples are usually invariant under a symmetry and the periodic orbits are degenerate. Therefore we discuss this case as well.

We assume that we have the following set-up. A compact Lie group G acts symplectically on a symplectic manifold (M, ω), i.e., for all $g \in G$ we have

$$g^*\omega = \omega.$$

Moreover, $H\colon M \to \mathbb{R}$ is a smooth function which is invariant under G, meaning that for all $g \in G$ we have

$$H \circ g = H.$$

Because the symplectic form as well as the function are invariant under G it follows that the Hamiltonian vector field is invariant as well. In other words, for all $g \in G$ we have

$$g^* X_H = X_H.$$

Therefore, if γ is a parametrized periodic orbit on the level set $H^{-1}(0)$ the same is true for $g \circ \gamma$ for every group element $g \in G$. Note that $g \circ \gamma$ has the same period as γ. To determine the kernel of the Hessian for a parametrized periodic orbit in the presence of a symmetry we abbreviate

$$\mathfrak{g} := \mathrm{Lie}(G)$$

the Lie algebra of the Lie group G. For $\xi \in \mathfrak{g}$ we denote by $X_\xi \in \Gamma(TM)$ the infinitesimal generator, namely the vector field given at $p \in M$ by

$$X_\xi(p) := \frac{d}{dt}\bigg|_{t=0} \exp(t\xi)(p).$$

This gives rise to a linear map

$$K_p \colon \mathfrak{g} \to T_pM, \quad \xi \mapsto X_\xi(p).$$

Note that if γ is a parametrized periodic orbit, then

$$\mathrm{im}(K_{\gamma(0)}) \times \{0\} \subset \mathfrak{K}(\gamma).$$

If we think of the symplectic action of G on (M, ω) as part of the structure and refer to the triple (M, ω, G) as a symplectic G-manifold, then we give a different notion of non-degeneracy for periodic orbits which does not coincide with the non-degeneracy condition when one forgets about the G-action.

Definition 7.3.3. A parametrized periodic orbit γ on a symplectic G-manifold is called *non-degenerate*, if

$$\mathfrak{K}(\gamma) = \langle (X_H(\gamma(0)), 0) \rangle + \big(\mathrm{im}(K_{\gamma(0)}) \times \{0\}\big).$$

A periodic orbit is non-degenerate if all of its parametrized representatives are non-degenerate.

Morally, one can think of a periodic orbit on a symplectic G-manifold as a periodic orbit on the quotient $H^{-1}(0)/G$ although this quotient is in general not a manifold. If one thinks of the periodic orbit as a periodic orbit in the quotient, the new notion of non-degeneracy coincides with the old one.

Again non-degeneracy on a G-manifold can be thought of as a Morse–Bott condition for the action functional $\mathcal{A}^H$. Because the reparametrization action of S^1 on the domain of the loop and the action of G on the target of the loop commute, we get a $G \times S^1$-action on the free loop space under which $\mathcal{A}^H$ is invariant. If a parametrized periodic orbit is non-degenerate this implies that its $G \times S^1$ orbit is a Morse–Bott component diffeomorphic to $G \times S^1$ modulo the stabilizer subgroup.

7.4 Periodic orbits of the first and second kind

In this section we discuss periodic orbits in the presence of an additional circle symmetry. This discussion will be relevant for the rotating Kepler problem where angular momentum generates a circle action, namely the rotation, under which the Hamiltonian is invariant. This phenomenon led Poincaré to his distinction between periodic orbits of the first and second kind. In this section we use the language of moment maps to introduce this dichotomy in a more general framework in which we address the question of non-degeneracy.

Assume that (M, λ) is a four-dimensional connected exact symplectic manifold with symplectic form $\omega = d\lambda$, which is endowed with a Hamiltonian action of the two-dimensional torus T^2, i.e., there exists a moment map

$$\mu = (\mu_1, \mu_2) \colon M \to \mathbb{R}^2$$

with the property that the Hamiltonian vector fields of the functions μ_1 and μ_2 commute, i.e.,

$$[X_{\mu_1}, X_{\mu_2}] = 0$$

and their flows are one-periodic, i.e.,

$$\phi_{\mu_k}^{t+1} = \phi_{\mu_k}^t, \quad t \in \mathbb{R}, \ k \in \{1, 2\}.$$

If $(r_1, r_2) \in T^2 = \mathbb{R}^2/\mathbb{Z}^2$ then its action is given by

$$(r_1, r_2)_*(p) = \phi_{\mu_1}^{r_1} \phi_{\mu_2}^{r_2}(p), \quad p \in M.$$

We assume that the action of the torus T^2 on M is *effective*. This means that for every $g \neq e = (0, 0) \in T^2$ there exists $p \in M$ such that $g_* p \neq p$, i.e., g moves at least one point. Alternatively this property might be expressed as follows. Denote for $p \in M$ by

$$G_p := \{g \in T^2 : g_* p = p\}$$

the isotropy group of p. Then the requirement that the action of T^2 is effective is equivalent to

$$\bigcap_{p \in M} G_p = \{e\}.$$

Although by definition each $g \neq e$ of an effective action moves only at least one point it actually moves a dense set of points as the next lemma shows.

Lemma 7.4.1. *Assume that the action of T^2 on M is effective. Then for each $g \neq e \in T^2$ the set $\{p \in M : g_* p \neq p\}$ is dense in M.*

Proof. Fix $g \neq e \in T^2$. We argue by contradiction and assume that there exists a nonempty open subset U of M with the property that $g_* p = p$ for every $p \in U$, i.e.,

$$U \subset \mathrm{Fix}(g) \tag{7.12}$$

where $\mathrm{Fix}(g)$ denotes the fixed point set of g. Abbreviate

$$T_g := \overline{\{g^n : n \in \mathbb{Z}\}}$$

the closure of the subgroup of T^2 generated by g. Note that the closure of a subgroup of a Lie group is itself a group and therefore a Lie subgroup of the given Lie group, because it is closed. We conclude that $T_g \subset T^2$ is a compact Lie subgroup of T^2. Therefore $\mathrm{Fix}(T_g)$ is a submanifold of M. In view of (7.12) it holds that

$$U \subset \mathrm{Fix}(T_g).$$

Consequently, $\mathrm{Fix}(T_g)$ has a four-dimensional connected component. Since M is connected this implies that $\mathrm{Fix}(T_g) = M$ contradicting the assumption that the action of T^2 on M is effective. This finishes the proof of the lemma. □

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. We abbreviate by

$$\mu_f\colon M \to \mathbb{R}$$

the smooth function

$$\mu_f := f \circ \mu.$$

Note that the Hamiltonian vector field of μ_f is given by

$$X_{\mu_f} = \partial_1 f X_{\mu_1} + \partial_2 f X_{\mu_2} \tag{7.13}$$

where $\partial_i f$ denotes the partial derivative of f with respect to the ith variable where $i \in \{1, 2\}$. We make the following standing hypothesis

(H1) 0 is a regular value of μ_f.

Note that it follows from (7.13) that if $\gamma \in C^\infty(S^1, \mu_f^{-1}(0))$ is a parametrized periodic orbit of X_{μ_f}, then $\mu \circ \gamma$ is constant, and hence only depends on the trace of γ. This justifies the following definition.

Definition 7.4.2. A periodic orbit is called of the *first kind* if it lies in the singular set of μ and of the *second kind* if it lies in the regular set of μ.

We first discuss periodic orbits of the first kind. To do that we introduce some notation. If $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ we abbreviate by $\mu_\xi\colon M \to \mathbb{R}$ the function

$$\mu_\xi = \xi_1\mu_1 + \xi_2\mu_2.$$

This agrees with the notation μ_f if we identify $\mathbb{R}^2$ with its dual space via the standard inner product and interpret ξ as an element in the dual space. A point $p \in \mu_f^{-1}(0)$ lies in the singular set of μ if and only if there exists $\xi \neq 0 \in \mathbb{R}^2$ such that p is a critical point of μ_ξ. Equivalently,

$$\xi_1 X_{\mu_1}(p) + \xi_2 X_{\mu_2}(p) = 0,$$

i.e., X_{μ_1} and X_{μ_2} are parallel. Because $p \in \mu_f^{-1}(0)$ and 0 is a regular value of μ_f it holds that

$$0 \neq X_{\mu_f}(p) = \partial_1 f(\mu(p)) X_{\mu_1}(p) + \partial_2 f(\mu(p)) X_{\mu_2}(p) \in \langle X_{\mu_1}(p) \rangle.$$

We conclude that p lies on a periodic orbit of X_{μ_f} on $\mu_f^{-1}(0)$. We have proved the following result

Lemma 7.4.3. *Through every point $p \in \mu_f^{-1}(0)$ which lies in the singular set of μ passes a periodic orbit of the first kind.*

We next discuss periodic orbits of the second kind. If p is a regular point of μ the isotropy group G_p is locally constant around p and because the action of T^2 on M is assumed to be effective we conclude by Lemma 7.4.1 that the isotropy group is trivial, i.e., $G_p = \{e\}$. Therefore in view of (7.13) the following result holds.

Lemma 7.4.4. *Through a point $p \in \mu_f^{-1}(0)$ in the regular set of μ a periodic orbit of the second kind passes, if and only if*

$$\frac{\partial_1 f(\mu(p))}{\partial_2 f(\mu(p))} \in \mathbb{Q}. \tag{7.14}$$

The minimal period $\tau_0 \in (0, \infty)$ of this orbit is determined by the condition

$$\partial_1 f(\mu(p))\tau_0, \partial_2 f(\mu(p))\tau_0 \in \mathbb{Z}, \quad \gcd\Big(\partial_1 f(\mu(p))\tau_0, \partial_2 f(\mu(p))\tau_0\Big) = 1$$

where gcd *denotes greatest common divisor.*

If at a point p in the singular set of μ condition (7.14) holds, periodic orbits of the second kind can bifurcate out from a periodic orbit of the first kind. In particular, the periodic orbit is then degenerate. Amazingly this is the only way how a periodic orbit of the first kind can become degenerate as the next theorem shows. Here we point out that non-degeneracy for a periodic orbit of the first kind refers to the non-degeneracy condition from Definition 7.3.2 and does not involve the torus action on the manifold M. Indeed, because the action of the torus is degenerate at p, the periodic orbit passing through p is fixed under the torus action and therefore isolated.

Theorem 7.4.5. *Assume that for $\xi \neq 0 \in \mathbb{R}^2$ a point $p \in \mathrm{crit}\mu_\xi \cap \mu_f^{-1}(0)$ satisfies*

$$\frac{\partial_1 f(\mu(p))}{\partial_2 f(\mu(p))} \notin \mathbb{Q}. \tag{7.15}$$

Then the periodic orbit determined by p is non-degenerate.

Proof. Note that because $p \in \mu_f^{-1}(0)$ and 0 is by (H1) a regular value of μ_f the Hamiltonian vector fields X_{μ_1} and X_{μ_2} cannot vanish simultaneously at p. In view of the assumption that the action of T^2 on M is effective it follows from the equivariant Darboux theorem due to Weinstein [239], see [73, Lemma 2.5], that there exists an equivariant neighborhood U of p in M with the following property. There exists $\epsilon > 0$ such that if $I = (-\epsilon, \epsilon)$ and $B = \{z \in \mathbb{C} : ||z|| < \epsilon\}$ there exists a diffeomorphism

$$\Psi \colon I \times S^1 \times B \to U$$

which meets the following conditions

(i) $\Psi(0, 0, 0) = p$.

(ii) $\Psi^* \omega|_U = \omega_0$ where ω_0 is the symplectic form

$$\omega_0 = da \wedge d\theta + dx \wedge dy$$

for coordinates $a \in (-\epsilon, \epsilon)$, $\theta \in S^1 = \mathbb{R}/\mathbb{Z}$ and $z = x + iy \in B \subset \mathbb{C}$.

(iii) If the torus T^2 acts on $I \times S^1 \times B$ by

$$(s_1, s_2)_*(a, \theta, z) = (a, \theta + s_1, e^{2\pi i s_2} z), \tag{7.16}$$

then there exists

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in SL_2(\mathbb{Z}),$$

i.e., an invertible 2×2-matrix with integer entries and determinant 1, such that if $\phi_A \colon T^2 \to T^2$ is the isomorphism of the torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ induced by A, then the map Ψ is equivariant in the sense that

$$\Psi\big(\phi_A(g)_*(y)\big) = g_*\big(\Psi(y)\big), \quad g \in T^2,\ y \in I \times S^1 \times B.$$

These conditions imply that

$$\mu \circ \Psi(a, \theta, z) = A \begin{pmatrix} a \\ \pi|z|^2 \end{pmatrix} + \mu(p).$$

We abbreviate

$$h \colon \mathbb{R}^2 \to \mathbb{R}, \quad \nu \mapsto f(A\nu + \mu(p))$$

and

$$\nu \colon I \times S^1 \times B \to \mathbb{R}^2, \quad (a, \theta, z) \mapsto (a, \pi|z|^2).$$

Note that ν is a moment map for the action of the torus (7.16) on $I \times S^1 \times B$ with respect to the symplectic form ω_0. It follows that

$$\mu_f \circ \Psi = h \circ \nu =: \nu_h. \tag{7.17}$$

The flow of the Hamiltonian vector field of ν_h is given by

$$\phi^t_{\nu_h}(a, \theta, z) = \big(a, \theta + t\partial_1 h(\nu), e^{2\pi i t \partial_2 h(\nu)} z\big). \tag{7.18}$$

Its differential at the origin becomes

$$d\phi^t_{\nu_h}(0)(\widehat{a},\widehat{\theta},\widehat{z}) = \big(\widehat{a},\widehat{\theta} + t\partial^2_{11}h(0)\widehat{a}, e^{2\pi i t\partial_2 h(0)}\widehat{z}\big).$$

Because

$$X_{\nu_h}(0) = \partial_1 h(0)\partial_\theta \in T_0 S^1$$

and

$$\ker d\nu_h(0) = T_0 S^1 \times T_0 B$$

we can identify

$$\ker d\nu_h(0)/\ker(\omega) = T_0 B = \mathbb{C}.$$

From (7.18) we read off that the minimal period of the periodic orbit of X_{ν_h} through the origin is

$$\tau_0 = \frac{1}{|\partial_1 h(0)|}.$$

For $n \in \mathbb{N}$ the induced map

$$\overline{d\phi^{n\tau_0}_{\nu_h}}(0)\colon \mathbb{C} \to \mathbb{C}$$

becomes

$$\overline{d\phi^{n\tau_0}_{\nu_h}}(0)\widehat{z} = e^{2\pi i n\tau_0\partial_2 h(0)}\widehat{z} = e^{2\pi i n\frac{\partial_2 h(0)}{|\partial_1 h(0)|}}\widehat{z}. \tag{7.19}$$

Assumption (7.15) can be rephrased as

$$\nabla f(\mu(p)) \notin \mathbb{R}\mathbb{Z}^2,$$

i.e., $\nabla f(\mu(p)$ is not of the form$(n_1 r, n_2 r)$ for $n_1, n_2 \in \mathbb{Z}$ and $r \in \mathbb{R}$. Because $A \in SL_2(\mathbb{Z})$, this is equivalent to the requirement that

$$\nabla h(0) \notin \mathbb{R}\mathbb{Z}^2,$$

respectively

$$\frac{\partial_1 h(0)}{\partial_2 h(0)} \notin \mathbb{Q}.$$

In particular, in view of (7.19) the linear symplectic map $\overline{d\phi^{n\tau_0}_{\nu_h}}(0)$ has no eigenvalue equal to one for every $n \in \mathbb{N}$. Therefore, by (7.11) the periodic orbit of X_{ν_h} passing through the origin is non-degenerate. By (7.17) the same is true for the periodic orbit of X_{μ_f} passing through p. This finishes the proof of the theorem. $\square$

Periodic orbits of the second kind are never isolated. Indeed, the torus T^2 acts on them and in the regular set of μ this action is free. However, we can ask if periodic orbits of the second kind are non-degenerate as periodic orbits of the symplectic T^2-manifold (M, ω, T^2) as in Definition 7.3.3. Note that if γ is a parametrized periodic orbit of the second kind, it follows from (7.13) that

$$(X_H(\gamma(0)), 0) \in \operatorname{im} K_{\gamma(0)}\mathfrak{t} \times \{0\}$$

where $\mathfrak{t} \cong \mathbb{R}^2$ is the Lie algebra of the torus T^2. Therefore γ is non-degenerate in the sense of Definition 7.3.3 if and only if

$$\mathfrak{K}(\gamma) = K_{\gamma(0)}\mathfrak{t} \times \{0\}. \tag{7.20}$$

We next describe a condition on the function f which guarantees that periodic orbits of the second kind are non-degenerate as periodic orbits on the T^2-manifold M. For $\mu \in \mathbb{R}^2$ we denote by

$$H_f(\mu)\colon \mathbb{R}^2 \to \mathbb{R}^2$$

the Hessian of f at μ with respect to the standard inner product on $\mathbb{R}^2$. By

$$\nabla f^{\perp} = \begin{pmatrix} -\partial_2 f \\ \partial_1 f \end{pmatrix}$$

we denote the skew gradient of f, i.e., the gradient of f rotated by 90 degrees. We make the following hypothesis on the function f.

(H2) $\langle \nabla f^{\perp}, H_f \nabla f^{\perp} \rangle \neq 0$ at every $\mu \in \mathbb{R}^2$.

Hypothesis (H2) might be equivalently rephrased without reference to the standard metric on $\mathbb{R}^2$ by the requirement that f has no critical point, i.e., $df(\mu) \neq 0$ for every $\mu \in \mathbb{R}^2$, and the quadratic form on $\ker df(\mu)$

$$v \mapsto d^2 f(\mu)(v, v), \quad v \in \ker df(\mu)$$

is non-degenerate for every $\mu \in \mathbb{R}^2$.

Example 7.4.6. An example of a function f satisfying hypothesis (H2) is the following. Let $h\colon \mathbb{R} \to \mathbb{R}$ be a smooth function with the property that its second derivative never vanishes, i.e.,

$$h''(\mu_1) \neq 0, \quad \mu_1 \in \mathbb{R}.$$

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the function

$$f(\mu_1, \mu_2) = h(\mu_1) + \mu_2, \quad (\mu_1, \mu_2) \in \mathbb{R}^2.$$

We compute the Hessian of f and find

$$H_f(\mu_1, \mu_2) = \begin{pmatrix} h''(\mu_1) & 0 \\ 0 & 0 \end{pmatrix}.$$

The skew gradient of f is given by

$$\nabla f^{\perp}(\mu_1, \mu_2) = \begin{pmatrix} -1 \\ h'(\mu_1) \end{pmatrix}.$$

Hence we obtain

$$\langle \nabla f^{\perp}, H_f \nabla f^{\perp} \rangle = h'' \neq 0.$$

Theorem 7.4.7. *Assume that f satisfies* (H2). *Then each periodic orbit of the second kind is non-degenerate in the sense of Definition* 7.3.3.

Proof. Assume that γ is a parametrized periodic orbit of the second kind of period τ with $p = \gamma(0)$. Since $K_p(\mathfrak{t})$ is a two-dimensional vector space and $T_p\mu_f^{-1}(0)$ is a three-dimensional vector space, it suffices to find a vector $v \in T_p\mu_f^{-1}(0)$ which has the property that $(v, \widehat{\tau}) \notin \mathfrak{K}(\gamma)$ for every $\tau \in \mathbb{R}$.

Because p is a regular point of μ, we can choose vectors $v_1, v_2 \in T_pM$ with the property that

$$d\mu_i(p)v_j = \delta_{ij}, \quad 1 \leq i, j \leq 2$$

where δ_{ij} is the Kronecker Delta. Define

$$v := -\partial_2 f(\mu(p))v_1 + \partial_1 f(\mu(p))v_2 \in T_pM.$$

We first compute

$$d\mu_f(p)v = \partial_1 f d\mu_1(v) + \partial_2 f d\mu_2(v) = -\partial_1 f \partial_2 f + \partial_1 f \partial_2 f = 0.$$

Moreover, we have

$$p = \phi^\tau_{\mu_f}(p) = \phi^{\partial_1 f(\mu(p))\tau}_{\mu_1}\phi^{\partial_2 f(\mu(p))\tau}_{\mu_2}.$$

We compute

$$\begin{aligned} d\phi^\tau_{\mu_f}(p)v = v &+ \Big(\partial_{11} f d\mu_1(v) + \partial_{12} f d\mu_2(v)\Big)X_{\mu_1} \\ &+ \Big(\partial_{21} f d\mu_1(v) + \partial_{22} f d\mu_2(v)\Big)X_{\mu_2} \\ = v &+ \Big(-\partial_{11} f \partial_2 f + \partial_{12} f \partial_1 f\Big)X_{\mu_1} + \Big(-\partial_{21} f\partial_2 f + \partial_{22} f \partial_1 f\Big)X_{\mu_2}. \end{aligned}$$

For the periodic orbit $\gamma \in C^\infty(S^1, \mu_f^{-1}(0))$ it holds that

$$\gamma(t) = \phi^{t\tau}_{\mu_f}(p), \quad t \in S^1.$$

Assume by contradiction that there exists $\widehat{\tau} \in \mathbb{R}$ such that

$$(v, \widehat{\tau}) \in \mathfrak{K}(\gamma).$$

In view of (7.13) this implies that

$$\begin{aligned} &\Big(-\partial_{11} f \partial_2 f + \partial_{12} f \partial_1 f\Big)X_{\mu_1} + \Big(-\partial_{21} f\partial_2 f + \partial_{22} f \partial_1 f\Big)X_{\mu_2} \\ &\quad = -\widehat{\tau}\Big(\partial_1 f X_{\mu_1} + \partial_2 f X_{\mu_2}\Big). \end{aligned}$$

Because p is a regular value of μ, the vector fields X_{μ_1} and X_{μ_2} are linearly independent at $\mu(p)$ so that we obtain

$$H_f \nabla f^\perp = -\widehat{\tau}\nabla f$$

implying

$$\langle \nabla f^\perp, H_f \nabla f^\perp \rangle = 0$$

in contradiction to hypothesis (H2). This proves the theorem. □

7.5 Symmetric periodic orbits and brake orbits

Suppose that (M,ω,ρ) is a real symplectic manifold, namely a symplectic manifold (M,ω) together with an anti-symplectic involution ρ. Moreover, assume that $H \in C^\infty(M,\mathbb{R})$ is a Hamiltonian invariant under ρ, i.e.,

$$H \circ \rho = H$$

and 0 is a regular value of H. Because the Hamiltonian is invariant under ρ and the symplectic form is anti-invariant, it follows that the Hamiltonian vector field is anti-invariant as well

$$\rho^* X_H = -X_H.$$

The anti-symplectic involution induces an involution $\mathcal{I}_\rho$ on the free loop space $C^\infty(S^1,\Sigma)$ of the energy hypersurface $\Sigma = H^{-1}(0)$. For a loop $\gamma \in C^\infty(S^1,\Sigma)$ this involution is given by

$$\mathcal{I}_\rho(\gamma)(t) = \rho(\gamma(1-t)), \quad t \in S^1.$$

Because the Hamiltonian vector field is anti-invariant under ρ, it follows that if γ is a periodic orbit of period τ, then $\mathcal{I}_\rho(\gamma)$ is again a periodic orbit of the same period τ. In particular, the involution $\mathcal{I}_\rho$ restricts to an involution of periodic orbits. A parametrized periodic orbit fixed under the involution $\mathcal{I}_\rho$ is called a *parametrized symmetric periodic orbit*. In particular, a parametrized symmetric periodic orbit satisfies

$$\gamma(t) = \rho(\gamma(1-t)).$$

If we plug into this equation $t = \frac{1}{2}$, we obtain

$$\gamma\big(\tfrac{1}{2}\big) = \rho\big(\gamma\big(\tfrac{1}{2}\big)\big)$$

concluding that

$$\gamma\big(\tfrac{1}{2}\big) \in \mathrm{Fix}(\rho) \cap \Sigma.$$

Because γ is periodic, i.e., $\gamma(0) = \gamma(1)$ we further obtain

$$\gamma(0) \in \mathrm{Fix}(\rho) \cap \Sigma \tag{7.21}$$

as well.

We discussed in Section 2.7 that the fixed point set of an anti-symplectic involution is a (maybe empty) Lagrangian submanifold of M. We further claim that

$$\Sigma \pitchfork \mathrm{Fix}(\rho), \tag{7.22}$$

meaning that Σ intersects the fixed point set of ρ transversely. To see that let $x \in \Sigma \cap \mathrm{Fix}(\rho) = H^{-1}(0) \cap \mathrm{Fix}(\rho)$. Because 0 is by assumption a regular value of H, there exists $v \in T_xM$ such that

$$dH(x)v \neq 0.$$

Because H is invariant under ρ, it holds that

$$dH(x) = d(H \circ \rho)(x) = dH(\rho(x))d\rho(x) = dH(x)d\rho(x).$$

Set

$$w := v + d\rho(x)v.$$

We compute

$$dH(x)w = dH(x)v + dH(x)d\rho(x)v = 2dH(x)v \neq 0.$$

Moreover,

$$d\rho(x)w = w$$

so that

$$w \in T_x\mathrm{Fix}(\rho).$$

This proves (7.22). We conclude that $\Sigma \cap \mathrm{Fix}(\rho)$ is an $n-1$-dimensional submanifold of the $2n-1$-dimensional manifold Σ if the dimension of M is $2n$.

We will call an unparametrized periodic orbit *symmetric* if the orbit can be parametrized as a symmetric periodic orbit. Note that not every parametrization of an unparametrized symmetric periodic orbit is a parametrized symmetric periodic orbit. Moreover, a symmetric parametrization of an unparametrized periodic orbit is not unique. Indeed, if γ is a parametrized symmetric periodic orbit, then the reparametrization $\left(\frac{1}{2}\right)_*\gamma = \gamma\left(\frac{1}{2} + \cdot\right)$ is a parametrized symmetric periodic orbit as well.

Lemma 7.5.1. *An unparametrized symmetric periodic orbit admits precisely two parametrizations as a parametrized symmetric periodic orbit.*

Proof. We first consider the simple case and assume that γ is a simple parametrized symmetric periodic orbit. Suppose that $r_*\gamma$ is also a parametrized symmetric periodic orbit for $r \in S^1 \setminus \{0\}$. From (7.21) we conclude that

$$r_*\gamma(0) = \gamma(r) \in \mathrm{Fix}(\rho).$$

Because $\rho^* X_H = -X_H$, it follows from uniqueness of solutions of an ODE that

$$\gamma(r+t) = \rho(\gamma(r-t)).$$

In particular,

$$\gamma(2r) = \rho(\gamma(0)) = \gamma(0),$$

where the second equality follows again from (7.21). Because γ is simple, we conclude

$$r = \tfrac{1}{2}.$$

Moreover, using again the fact that γ is simple we conclude

$$\left(\tfrac{1}{2}\right)_*\gamma \neq \gamma.$$

Therefore in the simple case the set

$$S_{[\gamma]} := \Big\{\gamma, \big(\tfrac{1}{2}\big)_* \gamma\Big\}$$

consists of all symmetric parametrized representatives of the unparametrized symmetric periodic orbit $[\gamma]$. If γ is not simple, then by looking at the underlying simple orbit, we conclude that in this case the set of symmetric parametrized periodic representatives is given by

$$S_{[\gamma]} = \Big\{\gamma, \big(\tfrac{1}{2\mathrm{cov}(\gamma)}\big)_* \gamma\Big\}.$$

This finishes the proof of the lemma. $\square$

The proof of the lemma actually reveals another interesting fact which we state as a separate lemma.

Lemma 7.5.2. *A parametrized symmetric periodic orbit γ intersects the fixed point set of ρ in precisely two points, namely*

$$\mathrm{im}(\gamma) \cap \mathrm{Fix}(\rho) = \Big\{\gamma(0), \gamma\big(\tfrac{1}{2\mathrm{cov}(\gamma)}\big)\Big\}.$$

If we restrict a parametrized symmetric periodic orbit γ of period τ to $[0, \frac{1}{2}]$, then the path $\gamma|_{[0,\frac{1}{2}]} \in C^\infty\big([0, \frac{1}{2}], \Sigma\big)$ is a solution of the problem

$$\begin{cases} \partial_t \gamma(t) = \tau X_H(\gamma(t)), & t \in \big[0, \frac{1}{2}\big] \\ \gamma(0),\ \gamma\big(\frac{1}{2}\big) \in \mathrm{Fix}(\rho). \end{cases}$$

A solution of this problem is referred to as a *brake orbit*. On the other hand given a brake orbit we can obtain a symmetric periodic orbit as follows. Namely we set for $t \in \big(\frac{1}{2}, 1\big]$

$$\gamma(t) := \rho(\gamma(1-t)).$$

In view of the boundary conditions of a brake orbit we obtain a continuous loop $\gamma \in C^0(S^1, \Sigma)$. Because X_H is anti-invariant under ρ and γ is an integral curve of the vector field τX_H on the interval $\big[0, \frac{1}{2}\big]$, it follows that γ is an integral curve of τX_H on the whole circle. In this case γ is automatically smooth and therefore a parametrized periodic orbit. Moreover, by construction it is symmetric. This proves that the restriction map

$$\gamma \mapsto \gamma|_{[0,\frac{1}{2}]}$$

induces a one-to-one correspondence between parametrized symmetric periodic orbits and brake orbits. We refer to τ as the period of the brake orbit. In particular, in our convention the period of the brake orbit coincides with the period of the corresponding parametrized symmetric periodic orbit.

Brake orbits are an interesting topic of study in their own right. The notion of brake orbits goes back to the work by Seifert [218], in which they were studied for mechanical Hamiltonians with anti-symplectic involution mapping p to $-p$ which corresponds to time reversal. We refer to the paper by Long, Zhang, and Zhu [169] for a modern study of brake orbits and as guide to the literature. We also mention the paper by Kang [142] in which brake orbits are studied in connection with respect to the restricted three-body problem.

There is a direct variational approach to brake orbits. Again to simplify the discussion we assume now in addition that (M, λ, ρ) is an exact real symplectic manifold, i.e., $\omega = d\lambda$ is a symplectic structure on M and ρ is a smooth involution on M satisfying $\rho^*\lambda = -\lambda$, which implies that $\rho^*\omega = -\omega$, i.e., ρ is an exact anti-symplectic involution. We observed in Section 2.7 of Chapter 2.1 that the fixed point set

$$L := \mathrm{Fix}(\rho)$$

is an exact Lagrangian submanifold of M. We introduce the space of paths

$$\mathfrak{P} := \Big\{ \gamma \in C^\infty\big(\big[0, \tfrac{1}{2}\big]\big) : \gamma(0), \gamma\big(\tfrac{1}{2}\big) \in L \Big\}$$

and consider the action functional

$$\mathcal{A}^H_\rho : \mathfrak{P} \times \mathbb{R}_+ \to \mathbb{R}$$

which is defined for $(\gamma, \tau) \in \mathfrak{P} \times \mathbb{R}_+$ by

$$\mathcal{A}^H_\rho(\gamma, \tau) = \int_0^{\frac{1}{2}} \gamma^*\lambda - \tau \int_0^{\frac{1}{2}} H(\gamma(t))dt.$$

This looks almost the same as the formula for $\mathcal{A}^H$, but note that the two action functionals are defined on different domains. One might think of the action functional $\mathcal{A}^H_\rho$ as an open string analog of the action functional $\mathcal{A}^H$. In fact, a Lagrangian analog of Rabinowitz Floer homology can be constructed using this action functional as was shown by Merry [175]. A computation as in the proof of Lemma 7.2.2 shows

Lemma 7.5.3. *Critical points of $\mathcal{A}^H_\rho$ precisely consist of pairs (γ, τ) where γ is a brake orbit of X_H on Σ of period τ.*

We next discuss the kernel of the Hessian of $\mathcal{A}^H_\rho$ at a critical point (γ, τ). If $(\widehat{\gamma}, \widehat{\tau})$ lies in the kernel of the Hessian we introduce analogously to (7.4) the path $v\colon \big[0, \tfrac{1}{2}\big] \to T_{\gamma(0)}M$ by

$$v(t) = d\phi_H^{-\tau t}(\gamma(t))\widehat{\gamma}(t).$$

The pair $(v, \widehat{\tau})$ still is a solution of the problem (7.6) with the only difference that the interval $[0, 1]$ has to be replaced by the interval $\big[0, \tfrac{1}{2}\big]$. However, the twist condition (7.5) has to be replaced by another condition taking into account that the periodic boundary conditions are now replaced by Lagrangian boundary conditions. First observe that the tangent space of $\mathfrak{P}$ at γ is given by

$$T_\gamma \mathfrak{P} = \Big\{\widehat{\gamma} \in \Gamma\big([0, \tfrac{1}{2}], \gamma^* TM\big) : \widehat{\gamma}(0) \in T_{\gamma(0)}L,\ \widehat{\gamma}\big(\tfrac{1}{2}\big) \in T_{\gamma(\frac{1}{2})}L\Big\}.$$

Therefore

$$v\big(\tfrac{1}{2}\big) = d\phi_H^{-\frac{\tau}{2}}\Big(T_{\gamma(\frac{1}{2})}L\Big).$$

Abbreviate

$$L_0 := T_{\gamma(0)}L,\ L_1 := d\phi_H^{-\frac{\tau}{2}}\Big(T_{\gamma(\frac{1}{2})}L\Big) \subset T_{\gamma(0)}M.$$

Note that both L_0 and L_1 are Lagrangian subspaces of the symplectic vector space $T_{\gamma(0)}M$. Abbreviating as in the periodic case $v_0 = v(0) = \widehat{\gamma}(0)$ we get the identification of the kernel of the Hessian of $\mathcal{A}^H_\rho$ at (γ, τ) with

$$\mathfrak{K}(\gamma) = \Big\{(v_0, \widehat{\tau}) \in L_0 \times \mathbb{R} : dH(\gamma(0))v_0 = 0,\ v_0 + \tfrac{\widehat{\tau}}{2} X_H(\gamma(0)) \in L_1\Big\}.$$

If $\pi\colon T\Sigma \to T\Sigma/\ker(\omega)$ denotes the canonical projection we abbreviate

$$\overline{L}_i := \pi(L_i \cap T_{\gamma(0)}\Sigma) \subset T_{\gamma(0)}\Sigma/\ker(\omega), \quad i \in \{0, 1\}.$$

As we have seen in (7.22), the Lagrangian $L = \mathrm{Fix}(\rho)$ intersects the energy hypersurface $\Sigma = H^{-1}(0)$ transversely. Therefore $\overline{L}_0$ and $\overline{L}_1$ are Lagrangian subspaces in the symplectic vector space $T_{\gamma(0)}\Sigma/\ker(\omega)$. Now the map $(v_0, \widehat{\tau}) \mapsto \pi(v_0)$ gives rise to a linear isomorphism

$$\mathfrak{K}(\gamma) \cong \overline{L}_0 \cap \overline{L}_1.$$

We are now in position to define non-degeneracy for brake orbits.

Definition 7.5.4. A brake orbit γ is called *non-degenerate* if $\mathfrak{K}(\gamma) = \{0\}$ or equivalently $\overline{L}_0 \pitchfork \overline{L}_1$.

If γ is a non-degenerate brake orbit of period τ, then the critical point (γ, τ) of the action functional $\mathcal{A}^H_\rho$ is actually a Morse critical point. This is in contrast to the periodic case, where at a non-degenerate parametrized periodic orbit the action functional gives rise to a Morse–Bott critical component diffeomorphic to a circle. The geometric reason for that is the following. The action functional $\mathcal{A}^H$ is invariant under the circle action obtained by rotating the loop. However, the action functional $\mathcal{A}^H_\rho$ is not anymore invariant under this action.

We next discuss a relation between the two notions of non-degeneracy.

Lemma 7.5.5. *Assume that γ is a non-degenerate parametrized symmetric periodic orbit. Then the brake orbit $\gamma|_{[0,\frac{1}{2}]}$ is non-degenerate as well.*

Proof. Because $\rho^* X_H = -X_H$, the differential $d\rho$ of the anti-symplectic involution ρ induces an anti-symplectic bundle map

$$R\colon T\Sigma/\ker(\omega) \to T\Sigma/\ker(\omega)$$

of the symplectic vector bundle $T\Sigma/\ker(\omega)$ over Σ which is characterized by the following commutative diagram

$$\begin{tikzcd} T\Sigma \arrow[r, "d\rho"] \arrow[d, "\pi"] & T\Sigma \arrow[d, "\pi"] \\ T\Sigma/\ker(\omega) \arrow[r, "R"] & T\Sigma/\ker(\omega). \end{tikzcd}$$

Recall from (7.10) the bundle map $\overline{d\phi_H^t}\colon T\Sigma/\ker(\omega) \to T\Sigma/\ker(\omega)$ induced from $d\phi_H^t$. Consider the two linear anti-symplectic involutions

$$R_0, R_1\colon T_{\gamma(0)}\Sigma/\ker(\omega) \to T_{\gamma(0)}\Sigma/\ker(\omega)$$

of the symplectic vector space $T_{\gamma(0)}\Sigma/\ker\omega$ defined as

$$R_0 := R_{\gamma(0)}, \quad R_1 := \overline{d\phi_H^{-\tau/2}}\big(\gamma(\tfrac{\tau}{2})\big)R_{\gamma(\frac{\tau}{2})}\overline{d\phi_H^{\tau/2}}\big(\gamma(0)\big).$$

Using this notion we can characterize the two Lagrangian subspaces L_0 and L_1 of $T_{\gamma(0)}\Sigma/\ker(\omega)$ as

$$\overline{L}_0 = \mathrm{Fix}(R_0), \quad \overline{L}_1 = \mathrm{Fix}(R_1). \tag{7.23}$$

Because $\rho^* X_H = -X_H$, the flow of X_H satisfies the following relation with ρ

$$\phi_H^t = \rho\phi_H^{-t}\rho.$$

Differentiating this formula we obtain

$$\overline{d\phi_H^t} = R\overline{d\phi_H^{-t}}R. \tag{7.24}$$

Suppose that

$$v_0 \in \overline{L}_0 \cap \overline{L}_1.$$

In view of (7.23) this means that

$$Rv_0 = v_0 \tag{7.25}$$

and

$$\overline{d\phi_H^{-\tau/2}}R\overline{d\phi_H^{\tau/2}}v_0 = v_0. \tag{7.26}$$

Using (7.24), (7.25), and (7.26) we compute

$$\begin{aligned} \overline{d\phi_H^\tau}v_0 &= \overline{d\phi_H^{\tau/2}} \circ \overline{d\phi_H^{\tau/2}}v_0 \\ &= R\overline{d\phi_H^{-\tau/2}}R\overline{d\phi^{\tau/2}}v_0 \\ &= Rv_0 \\ &= v_0 \end{aligned}$$

implying that

$$v_0 \in \ker\left(\overline{d\phi_H^\tau} - \mathrm{id}\right),$$

i.e., v_0 is an eigenvalue of the linear map $\overline{d\phi_H^\tau}$ to the eigenvalue 1. Because the parametrized periodic orbit γ is non-degenerate by assumption, this implies in view of (7.11) that

$$v_0 = 0.$$

We have shown that

$$\overline{L}_0 \pitchfork \overline{L}_1$$

and therefore the brake orbit $\gamma|_{[0,\frac{1}{2}]}$ is non-degenerate. This finishes the proof of the lemma. □

We remark that there is no converse to Lemma 7.5.5. If γ is a parametrized symmetric periodic orbit with the property that the corresponding brake orbit $\gamma|_{[0,\frac{1}{2}]}$ is non-degenerate, we cannot infer that γ itself is non-degenerate. For example it might happen that non-symmetric parametrized periodic orbit bifurcates out of the symmetric parametrized periodic orbit γ. In this case γ is degenerate as a parametrized periodic orbit. However, there is no reason that $\gamma|_{[0,\frac{1}{2}]}$ should be degenerate as a break orbit as well, since the non-symmetric parametrized periodic orbits do not give rise to brake orbits.

We next explain how Lemma 7.5.5 generalizes to the case with symmetry. Suppose that G is a compact connected Lie group with Lie algebra $\mathfrak{g}$. Recall from Chapter 3, Section 3.2 that a *Hamiltonian action* of G on a symplectic manifold (M, ω) is a smooth action of G on M for which there exists a *moment map*, namely a smooth map

$$\mu\colon M \to \mathfrak{g}^*$$

where $\mathfrak{g}^*$ is the dual of the Lie algebra $\mathfrak{g}$ which meets the following two requirements

(i) μ is equivariant with respect to the action of G on M and the coadjoint action of G on $\mathfrak{g}^*$.

(ii) For each $\xi \in \mathfrak{g}$ the Hamiltonian vector field of the associated smooth function $\langle \mu, \xi \rangle\colon M \to \mathbb{R}$ satisfies

$$X_{\langle \mu, \xi \rangle} = X_\xi$$

where $X_\xi(p) = \frac{d}{dt}\big|_{t=0} \exp(t\xi)p$ for $p \in M$ is the infinitesimal generator.

Note that a Hamiltonian action is necessarily symplectic. If now (M, ω, ρ) is a real symplectic manifold, we say that a Hamiltonian action of G on M is *real* if the moment map is real in the sense that

$$\mu \circ \rho = \mu.$$

Observe that for a real Hamiltonian action it holds that

$$\rho^* X_\xi = -X_\xi, \quad \forall\, \xi \in \mathfrak{g}.$$

In particular, if $p \in L = \mathrm{Fix}(\rho)$, it holds that

$$T_pL \pitchfork K_p(\mathfrak{g}) \tag{7.27}$$

where we recall that $K_p \colon \mathfrak{g} \to T_pM$ is the linear map $\xi \mapsto X_\xi(p)$.

Remark 7.5.6. There is a more general notion of real moment map which needs an additional involution on the Lie group G, see [113, 196]. In our case the involution on G is just given by the map $g \mapsto g^{-1}$ for $g \in G$.

Using the transverse intersection in (7.27) the proof of Lemma 7.5.5 reveals the following generalization to the case with symmetry.

Lemma 7.5.7. *Assume that we have a real Hamiltonian action of the Lie group G on (M, ω, ρ) and γ is a parametrized symmetric periodic orbit which is non-degenerate on the G-manifold M in the sense of Definition* 7.3.3. *Then the brake orbit $\gamma|_{[0,\frac{1}{2}]}$ is non-degenerate as well.*

We finally discuss a special class of symmetric periodic orbits which are relevant for the restricted three-body problem. As in Section 6.3 in Chapter 6 let

$$I \colon S^2 \to S^2, \quad (x_1, x_2, x_3) \mapsto (-x_1, x_2, x_3)$$

be the reflection at the great arc

$$\{(0, x_2, x_3) : x_2^2 + x_3^3 = 1\} \subset \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 = 1\} = S^2$$

and

$$\rho \colon T^*S^2 \to T^*S^2$$

be the anti-symplectic involution obtained by composing the map d_*I with the anti-symplectic involution $(p, q) \mapsto (p, -q)$ on T^*S^2. Here we use the nonstandard convention motivated by Moser regularization to denote points on S^2 by p and cotangent vectors by q. Note that the fixed point set of ρ is the conormal bundle over the equator. Suppose that

$$\Sigma \subset T^*S^2$$

is a fiberwise starshaped hypersurface which is invariant under ρ, i.e.,

$$\rho(\Sigma) = \Sigma.$$

Observe that $\Sigma \cap \mathrm{Fix}(\rho)$ is diffeomorphic to the unit conormal bundle over the equator of S^2, and hence consists of two circles. We write

$$\Sigma \cap \mathrm{Fix}(\rho) = S_1 \cup S_2$$

where S_1 and S_2 are the two connected components of $\Sigma \cap \mathrm{Fix}(\rho)$, both diffeomorphic to a circle. Note that Σ is diffeomorphic to $\mathbb{R}P^3$, so its fundamental group is $\pi_1(\Sigma) = \mathbb{Z}/2\mathbb{Z}$.

Proposition 7.5.8. *Assume that $\gamma\colon S^1 \to \Sigma$ is a parametrized symmetric periodic orbit. Then γ is contractible if and only if $\gamma(0)$ and $\gamma\left(\frac{1}{2}\right)$ lie in the same connected component of $\Sigma \cap \mathrm{Fix}(\rho)$.*

Proof. Consider the Levi-Civita map

$$\mathscr{L}\colon \mathbb{C}^2 \setminus \{0\} \to T^*S^2 \setminus S^2, \quad (u,v) \mapsto \left(\frac{u}{\overline{v}}, 2v^2\right).$$

On $\mathbb{C}^2$ endowed with the symplectic form $\omega_{\mathscr{L}} = 4(dv_1 \wedge du_1 + dv_2 \wedge du_2)$ we have the two anti-symplectic involutions $\rho_1, \rho_2\colon \mathbb{C}^2 \to \mathbb{C}^2$ given by

$$\rho_1(u,v) = (\overline{u}, -\overline{v}), \quad \rho_2(u,v) = (-\overline{u}, \overline{v}).$$

Note that

$$\rho \circ \mathscr{L} = \mathscr{L} \circ \rho_1, \quad \rho \circ \mathscr{L} = \mathscr{L} \circ \rho_2.$$

Moreover, the two anti-symplectic involutions on $\mathbb{C}^2$ commute with each other and their composition is the symplectic involution

$$\rho_1 \circ \rho_2 = \rho_2 \circ \rho_1 = -\mathrm{id}.$$

Note that the preimage of the fiberwise starshaped hypersurface Σ under the Levi-Civita map

$$\widetilde{\Sigma} := \mathscr{L}^{-1}(\Sigma) \subset \mathbb{C}^2 \setminus \{0\}$$

is a starshaped hypersurface in $\mathbb{C}^2$. Let

$$\widetilde{S}_1 := \mathscr{L}^{-1}(S_1), \quad \widetilde{S}_2 := \mathscr{L}^{-1}(S_2)$$

be the preimages of the two circles S_1 and S_2 in $\widetilde{\Sigma}$. Note that $\widetilde{S}_1$ and $\widetilde{S}_2$ are themselves circles which doubly cover the circles S_1 and S_2. Moreover, maybe after interchanging the indices of S_1 and S_2 we have

$$\widetilde{S}_1 = \mathrm{Fix}(\rho_1) \cap \widetilde{\Sigma}, \quad \widetilde{S}_2 = \mathrm{Fix}(\rho_2) \cap \widetilde{\Sigma}.$$

We first consider a parametrized symmetric periodic orbit $\gamma\colon S^1 \to \Sigma$ that is contractible. It follows that there exists a lift $\widetilde{\gamma}$ to $\widetilde{\Sigma}$ of γ which is itself periodic, i.e.,

$$\widetilde{\gamma}\colon S^1 \to \widetilde{\Sigma}, \quad \mathscr{L} \circ \widetilde{\gamma} = \gamma.$$

Because γ is symmetric, it holds that $\gamma(0) \in \Sigma \cap \mathrm{Fix}(\rho)$. We only discuss the case where

$$\gamma(0) \in S_1$$

the alternative case $\gamma(0) \in S_2$ is completely analogous. It follows that

$$\widetilde{\gamma}(0) \in \widetilde{S}_1.$$

Because γ is symmetric with respect to ρ, it follows that its lift satisfies

$$\widetilde{\gamma}(t) = \pm\rho_1(\widetilde{\gamma}(1-t)).$$

Since $\widetilde{\gamma}(0) \in \mathrm{Fix}(\rho_1)$, we conclude that

$$\widetilde{\gamma}(t) = \rho_1(\widetilde{\gamma}(1-t)).$$

This implies that

$$\widetilde{\gamma}\big(\tfrac{1}{2}\big) \in \mathrm{Fix}(\rho_1) \cap \widetilde{\Sigma} = \widetilde{S}_1.$$

Consequently

$$\gamma\big(\tfrac{1}{2}\big) \in S_1$$

and $\gamma(0)$ and $\gamma\big(\tfrac{1}{2}\big)$ lie in the same connected component of $\mathrm{Fix}(\rho) \cap \Sigma$.

It remains to show the converse implication. Namely we assume that $\gamma(0)$ and $\gamma\big(\tfrac{1}{2}\big)$ lie in the same connected component of $\Sigma \cap \mathrm{Fix}(\rho)$ and we show that γ is contractible. Again we only discuss the case where $\gamma(0) \in S_1$, the other one is completely analogous. We consider a lift

$$\widetilde{\gamma}\colon [0,1] \to \widetilde{\Sigma}, \quad \mathscr{L} \circ \widetilde{\gamma} = \gamma.$$

We have to show that the lift $\widetilde{\gamma}$ is a loop, i.e., $\widetilde{\gamma}(1) = \widetilde{\gamma}(0)$. Because $\gamma(0) \in S_1$, it holds that

$$\widetilde{\gamma}(0) \in \widetilde{S^1} = \mathrm{Fix}(\rho_1) \cap \widetilde{\Sigma}$$

and because γ is symmetric, we conclude that

$$\widetilde{\gamma}(t) = \rho_1(\widetilde{\gamma}(1-t)), \quad t \in [0,1].$$

In particular,

$$\widetilde{\gamma}(1) = \rho_1(\widetilde{\gamma}(0)) = \widetilde{\gamma}(0)$$

where the last equality follows from the fact that $\widetilde{\gamma}(0) \in \mathrm{Fix}(\rho_1)$. This proves that $\widetilde{\gamma}$ is a loop and the proposition follows. $\square$

7.6 Blue sky catastrophes

Assume that (M,ω) is a symplectic manifold and $H \in C^\infty(M \times [0,1], \mathbb{R})$. We think of H as a one parameter family of autonomous Hamiltonian functions $H_r = H(\cdot, r) \in C^\infty(M,\mathbb{R})$ and we assume that 0 is a regular value of H_r for every $r \in [0,1]$, the level set $H_r^{-1}(0)$ is connected for every $r \in [0,1]$, and $H^{-1}(0)$ is compact such that $H_r^{-1}(0)$ is a smooth one parameter family of closed, connected submanifolds of M. For $r \in [0,1]$ abbreviate by X_{H_r} the Hamiltonian vector field of H_r implicitly defined by the condition

$$dH_r = \omega(\cdot, X_{H_r}).$$

Suppose that $(\gamma_r, \tau_r) \in C^\infty\big(S^1, H_r^{-1}(0) \times (0,\infty)\big)$ for $r \in [0,1)$ is a smooth family of loops and positive numbers solving the problem

$$\partial_t \gamma_r(t) = \tau_r X_{H_r}(\gamma_r(t)), \quad t \in S^1, \ r \in [0,1),$$

i.e., γ_r is a smooth family of periodic orbits of X_{H_r} on the energy hypersurfaces $H_r^{-1}(0)$ and τ_r are its periods. Suppose now that τ_r converges to $\tau_1 \in (0,\infty)$ as r goes to 1. Because $H^{-1}(0)$ is compact it follows from the theorem of Arzelà–Ascoli that γ_r converges to a periodic orbit γ_1 of period τ_1. On the other hand if τ_r goes to infinity, then the family of periodic orbits γ_r "disappears in the blue sky" as r goes to 1. Such a scenario is referred to as a "*blue sky catastrophe*". The term goes back to Abraham, see [17, Chapter 3, Section 2.7], [200, Problem 37].

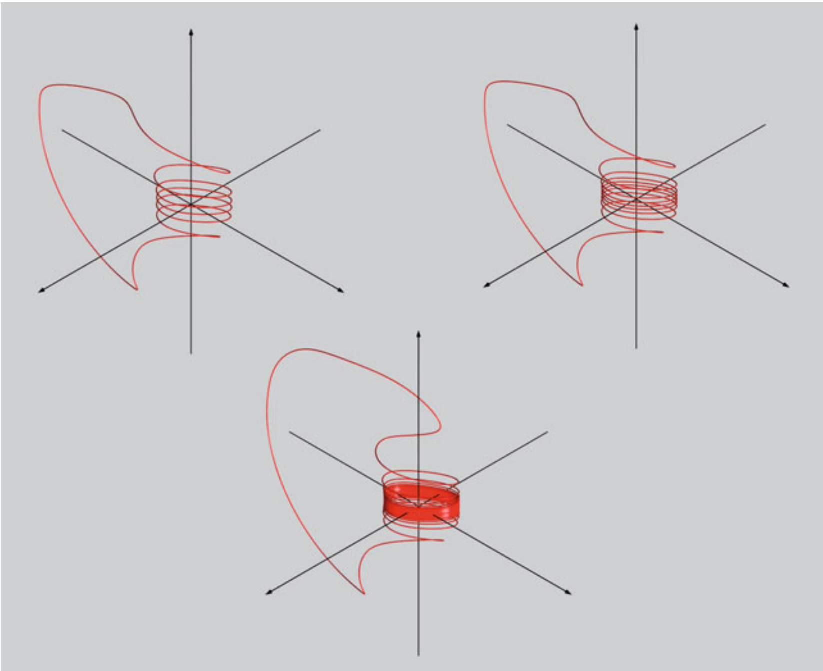


Figure 7.2: A blue sky catastrophe in the Gavrilov–Shilnikov model, [96]: the period of a periodic orbit blows up as one parameter converges to some value: this never happens for a Reeb flow on a compact manifold.

We explain how the assumption that the energy hypersurfaces $H_r^{-1}(0)$ are contact prevents blue sky catastrophes.

Theorem 7.6.1. *Assume that $\omega = d\lambda$ such that $\lambda|_{H_r^{-1}(0)}$ is a contact form for every $r \in [0,1]$. Suppose that (γ_r, τ_r) for $r \in [0,1)$ is a smooth family of periodic orbits γ_r of period τ_r. Then there exists $\tau_1 \in (0,\infty)$ such that τ_r converges to τ_1.*

Proof. Let R_r be the Reeb vector field of $\lambda|_{H_r^{-1}(0)}$. Note that $X_{H_r}|_{H_r^{-1}(0)}$ is parallel to R_r. We first claim that we can assume without loss of generality that

$$R_r = X_{H_r}|_{H_r^{-1}(0)}. \tag{7.28}$$

To see this note that because the two vector fields are parallel there exist smooth functions $f_r\colon H_r^{-1}(0) \to \mathbb{R}$ such that

$$R_r = f_r X_{H_r}|_{H_r^{-1}(0)}.$$

Moreover, because $H^{-1}(0)$ is compact, there exists $c > 0$ such that

$$\frac{1}{c} \leq |f_r(x)| \leq c, \quad r \in [0,1],\ x \in H_r^{-1}(0).$$

Now choose a smooth extension $\bar{f}\colon M \times [0,1] \to \mathbb{R} \setminus \{0\}$ such that

$$\bar{f}(\cdot, r)|_{H_r^{-1}(0)} = f_r$$

and replace H by $\bar{f} \cdot H$. This guarantees (7.28). The original family of periodic orbits γ_r gets reparametrized by this procedure. However, because of the compactness of $H^{-1}(0)$ the question about convergence of τ_r is unaffected.

We now consider the family of functionals

$$\mathcal{A}_r := \mathcal{A}^{H_r}\colon C^\infty(S^1, M) \times (0,\infty) \to \mathbb{R}.$$

Using (7.28) we compute now the action of $\mathcal{A}_r$ at the critical point (γ, τ) as follows

$$\mathcal{A}_r(\gamma,\tau) = \int_0^1 \lambda\big(\tau X_{H_r}(\gamma)\big)dt = \tau \int_0^1 \lambda(R_r)dt = \tau, \tag{7.29}$$

i.e., the period of the periodic orbit γ can be interpreted as the action value of $\mathcal{A}_r$. Now let (γ_r, τ_r) for $r \in [0,1)$ be a smooth family of periodic orbits, or according to our new interpretation critical points of $\mathcal{A}_r$. Using (7.29) we are now in position to compute the derivative of τ_r with respect to the r-parameter as follows

$$\partial_r \tau_r = \frac{d}{dr}\big(\mathcal{A}_r(\gamma_r,\tau_r)\big) = (\partial_r \mathcal{A}_r)(\gamma_r,\tau_r) = -\tau_r \int_{S^1} (\partial_r H_r)(\gamma_r)dt. \tag{7.30}$$

Here we have used in the second equation that (γ_r, τ_r) is a critical point of $\mathcal{A}_r$. Because $H^{-1}(0)$ is compact, there exists $\kappa > 0$ such that

$$\Big|\partial_r H_r|_{H_r^{-1}(0)}\Big| \leq \kappa, \quad \forall\ r \in [0,1]. \tag{7.31}$$

Combining (7.30) and (7.31) we obtain the estimate

$$|\partial_r \tau_r| \leq \kappa \tau_r.$$

In particular, if $0 \leq r_1 < r_2 < 1$, this implies

$$e^{-\kappa(r_2 - r_1)}\tau_{r_1} \leq \tau_{r_2} \leq e^{\kappa(r_2 - r_1)}\tau_{r_1}.$$

This proves that τ_r converges, when r goes to 1 and hence excludes blue sky catastrophes. $\square$

7.7 Elliptic and hyperbolic orbits

Suppose that (γ, τ) is a periodic orbit of X_H on the energy level $\Sigma = H^{-1}(c)$. We then choose a disk $D : D^{2n-2} \mapsto \Sigma^{2n-1}$ such that

- $\mathrm{im}(\gamma) \cap \mathrm{im}(D) = x_0$
- the disk is transverse to the flow, so $TD(D^{2n-2}) \oplus X_H(x_0) = T_{x_0}\Sigma$.

We will write $V := T_{x_0} D(D^{2n-2})$. The disk D is called a *local surface of section*. It can be used to translate certain questions about periodic orbits into the study of the return map, which we will define now. Choose a contractible, open subset $\tilde{D}^{2n-2} \subset D^{2n-2}$ such that for all $y \in \tilde{D}^{2n-2}$ there is a time $t > 0$ such that $\phi^t_{X_H} \circ D(y) \in D(D^{2n-2})$. Such a time exists for x_0 and hence for nearby points.

Given $y \in \tilde{D}^{2n-2}$, choose the smallest $t(y)$ such that $\phi^{t(y)}_{X_H} \circ D(y) \in D(D^{2n-2})$. By smooth dependence on initial conditions, $t(y)$ depends smoothly on y. The return map defined as the map

$$\phi : \tilde{D}^{2n-2} \longrightarrow D^{2n-2}, \quad D^{-1} \circ \phi^{t(y)}_{X_H} \circ D(y),$$

is therefore smooth. It is clear from the definition that there is a one-to-one correspondence between periodic points of ϕ and periodic orbits near (γ, τ). This is illustrated in Figure 7.3. In practice, the linearized return map, defined as the

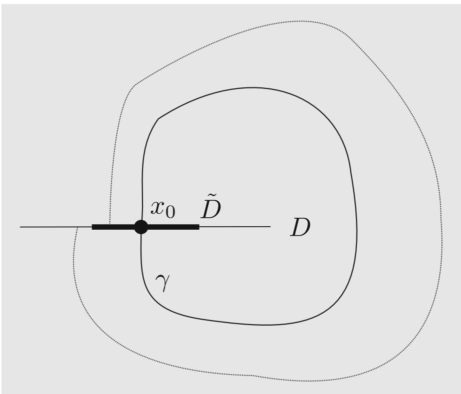


Figure 7.3: A local surface of section near a periodic orbit γ.

linear map

$$d_{x_0}\phi : V \to V,$$

plays an important role. Before we discuss this, we remind the reader that the linearized flow for time τ, given by $d_{x_0}\phi^\tau_{X_H}$ has always a non-trivial kernel because $X_H \in \ker d_{x_0}\phi^\tau_{X_H}$. Poincaré observed that the restriction of the flow to a hypersurface transverse to it, eliminates this degeneracy. This makes the linearized return map more useful, and further motivates the definition of a transversely non-degenerate, periodic orbit (γ, τ), which we have seen in Definition 7.3.1. The associated periodic point on the local surface of section is then also called *non-degenerate*.

In case of transversely non-degenerate periodic orbits, one can show that these orbits persist under perturbations of the vector field. This does not depend on the vector field being Hamiltonian, see for instance [189, Theorem 2.2]. A proof only needs the non-degeneracy, and an application of the implicit function theorem. Here is the statement.

Theorem 7.7.1. *Suppose X_μ is a smooth 1-parameter family of vector fields, and assume that (γ_0, τ_0) is a transversely non-degenerate periodic orbit of X_0. Then for μ sufficiently close to 0 there is a periodic orbit (γ_μ, τ_μ) with period τ_μ close to τ_0 such that*

$$\lim_{\mu \to 0} \tau_\mu = \tau, \quad \lim_{\mu \to 0} \gamma_\mu(t) = \gamma(t).$$

This periodic orbit (γ_μ, τ_μ) is unique up to a time-shift.

The Hamiltonian case has special features which are of interest to us. Indeed, the linearized return map, the map $d_{x_0}\phi$, is a linear symplectomorphism, and we recall from symplectic linear algebra that eigenvalues for such maps come in pairs.

Lemma 7.7.2. *Suppose A is a symplectic matrix in $Sp(n)$. Then the characteristic polynomial of a symplectic matrix A is "symmetric", i.e., there are $a_0, \ldots, a_n$ such that*

$$\Delta(\lambda) = \lambda^{2n} \Delta\left(\frac{1}{\lambda}\right) = \lambda^n \sum_{k=0}^{n} a_k (\lambda^k + \lambda^{-k}).$$

Furthermore if λ is an eigenvalue of A, then so are λ^{-1}, $\bar{\lambda}$ and $\bar{\lambda}^{-1}$.

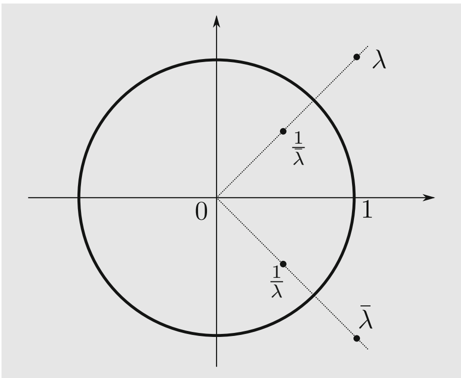


Figure 7.4: Distribution of eigenvalues of a symplectic matrix.

Proof. As a reminder, here is an argument. We start with the first identity. Since A is symplectic, we have $J_0^{-1}A^tJ_0 = A^{-1}$, so

$$\begin{aligned}\Delta(\lambda) &= \det A \det(\mathrm{id} - \lambda A^{-1}) = \det A \det J_0^{-1} \det(\mathrm{id} - \lambda A^t) \det J_0 \\ &= (-\lambda)^{2n} \det(A^t - \lambda^{-1}\,\mathrm{id}) = \lambda^{2n}\Delta\left(\frac{1}{\lambda}\right).\end{aligned}$$

Since A is real, so is the characteristic polynomial $\Delta(\lambda) = \sum_{k=0}^{2n} b_k\lambda^k$. By the first identity, $\sum_{k=0}^{2n} b_k\lambda^k = \sum_{k=0}^{2n} b_k\lambda^{2n-k}$, so we find $b_{n+k} = b_{n-k}$, which shows the second identity if we put

$$a_k = \begin{cases} \frac{b_n}{2} & k = 0, \\ b_{n+k} & k > 0. \end{cases}$$

Finally, if λ is a zero of Δ, we see that λ^{-1} is, too, by the symmetry property. As Δ is real, we find the statement for $\bar{\lambda}$ and $\bar{\lambda}^{-1}$ as well. □

In dimension 2, this implies that eigenvalues of a symplectic matrix are either both real or they lie on the unit circle.

Definition 7.7.3. Call a fixed point x_0 of a symplectomorphism ϕ *elliptic* if the eigenvalues of $d_{x_0}\phi$ lie on the unit circle. We call a fixed point *hyperbolic* if the eigenvalues of $d_{x_0}\phi$ lie on the real axis.

Given the correspondence between periodic points of the return map and periodic orbits, we call a periodic orbit of a flow elliptic or hyperbolic if its linearized return map is elliptic or hyperbolic. Geometrically, the linearized return map at an elliptic fixed point is conjugate to a rotation.

Chapter 8

Periodic Orbits in the Restricted Three-Body Problem

8.1 Some heroes in the search for periodic orbits

There are many heroes who devoted a great part of their life in order to find periodic orbits in the restricted three-body problem and thereby gave us precious insight into its intricate dynamics. We can only mention a few of them. The first hero we already met in Section 7.1 in Chapter 7. This is *Hill* who found power series for the direct and retrograde periodic orbit in Hill's lunar problem. The second hero is of course *Poincaré* who brought the importance of the search for periodic orbits to the fore. In his famous "Nouvelles méthodes de la mécanique céleste" [205] Poincaré distinguishes between periodic orbits of the first and second kind for small mass ratios μ. Periodic orbits of the first kind are orbits which bifurcate out of the circular periodic orbits of the rotating Kepler problem for $\mu = 0$ and periodic orbits of the second kind are the ones which bifurcate out of the elliptic periodic orbits of the rotating Kepler problem. Poincaré introduced also the notion of periodic orbits of the third kind in the spatial restricted three-body problem. These have non-vanishing inclination. However, since we restrict our attention to the planar case we do not discuss them further. The essential ideas how to prove existence of periodic orbits of the first and second kind are contained in [205]. However, the proof of existence of periodic orbits of the second kind is quite intricate. This is due to the fact that the elliptic periodic orbits in the rotating Kepler problem are degenerate. Rigorous proofs for the existence of periodic orbits of the second kind were found by *Arenstorf* and *Barrar* in [15, 25, 26].

The rotating Kepler problem is one source of known periodic orbits in the restricted three-body problem. Namely, one fixes an energy value and starts varying the mass ratio μ and then follows a family of periodic orbits which starts at one of the known periodic orbits of the rotating Kepler problem. Another approach to find periodic orbits in the restricted three-body problem is that one fixes μ and

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_8

follows a family by varying the energy. Of course one needs something to start with. One option is to start the family at a Lagrange point. Another possibility is to let the energy go to $-\infty$. Then for each bounded component one approaches the Kepler problem which after regularization becomes the geodesic flow on the round two-sphere. A parametrized simple geodesic on S^2 is prescribed by a starting point and a unit direction. Therefore the moduli space of parametrized simple geodesics is diffeomorphic to the unit tangent bundle of S^2 which itself is diffeomorphic to $\mathbb{R}P^3$. The moduli space of unparametrized simple periodic geodesics becomes then $\mathbb{R}P^3/S^1 \cong S^2$. The simplest Morse function on S^2 is the height function having one maximum and one minimum. It turns out that for very low energy periodic orbits of the restricted three-body problem bifurcate from the Kepler problem like the critical points of the height function on S^2, namely the periodic orbit corresponding to the maximum is the direct periodic orbit and the periodic orbit corresponding to the minimum is the retrograde periodic orbit.

The astronomer George *Darwin*, the son of Charles Darwin, was one of the pioneers in applying numerical methods to detect periodic orbits in the restricted three-body problem. He considered the case of $\mu = \frac{1}{11}$, i.e., the earth is ten times heavier than the moon, and mainly considers the direct periodic orbit around the moon, i.e., the lighter body, which Darwin calls Jupiter, see [68, 69]. Another of the early pioneers in the search for periodic orbits in the restricted three-body problem was *Moulton* and his school, whose results are summarized in [190]. The most complete classification of periodic orbits for the case $\mu = \frac{1}{2}$, i.e., the two primaries have equal mass, was carried out by *Strömgren* and his school at the Observatory in Copenhagen between 1913–1939, see [230]. Broucke [45] considered the case of earth and moon. For a nice summary of these numerical results we refer to Chapter 9 in the impressive book by *Szebehely* [231]. Nevertheless the numerical exploration of periodic orbits in the restricted three-body problem goes on. The books by *Bruno* [49] and *Hénon* [115] discuss more recent results to which the authors of these books greatly contributed. We refer the reader also to the survey article of Hénon's and Bruno's work by Bathkin and Bathkina [29].

Periodic orbits in Hill's lunar problem were investigated numerically by, for instance, *Matukuma* in [171] and Hénon in [114]. Their numerical results suggest that the simple and doubly covered retrograde periodic orbit is always nondegenerate while the simple direct periodic orbit becomes degenerate below the first critical value at which point two other periodic orbits bifurcate out of the direct one. Looking at the numerical results this seems to be a general phenomenon in the restricted three-body problem. The retrograde at least below the first critical value seems to be rather boring, while the direct one shows rather wild behavior, as already the pictures computed by Darwin vividly show.

A result contrasting sharply with the numerical results described above was obtained by Birkhoff in [39]. Using a shooting argument he proved existence of a retrograde periodic orbit for all mass ratios $\mu \in [0, 1)$ below the first critical value.

8.2 Periodic orbits in the rotating Kepler problem

Recall that the Hamiltonian of the rotating Kepler problem is given by

$$H = E + L$$

where E is the Hamiltonian of the (non-rotating) Kepler problem and L is angular momentum. Furthermore, by (3.14) the relation

$$|A|^2 = 1 + 2L^2 E$$

holds, where A is the Runge–Lenz vector whose length corresponds to the eccentricity of the corresponding Kepler ellipse. Combining these two facts we obtain the inequality

$$0 \leq 1 + 2E(H-E)^2 = 1 + 2H^2E - 4HE^2 + 2E^3 =: p(H, E). \tag{8.1}$$

Moreover, equality holds if and only if the corresponding Kepler orbit has vanishing eccentricity, i.e., for circular periodic orbits. Before we investigate these circular orbits in more detail, we first explain some symmetry properties that general orbits enjoy.

8.2.1 The shape of the orbits if $E < 0$

From Noether's theorem we know that $\{E, L\} = 0$, so $[X_E, X_L] = 0$. It follows that the flows of X_E and X_L commute. In particular, using the notation from (5.3) we see that

$$H = E + L = E \diamond L$$

so that

$$\phi_H^t = \phi_L^t \circ \phi_E^t. \tag{8.2}$$

We want to investigate what the orbits look like if $E < 0$.

For this we consider the q-components of an orbit in the Kepler problem. Let $\epsilon_\tau : [0, \tau] \to \mathbb{R}^2$ denote a Kepler ellipse with period τ, i.e., a solution to the Kepler problem with negative energy.

By (8.2), we also obtain a solution to the rotating Kepler problem, which no longer needs to be periodic. Its q-components are given by

$$\epsilon_\tau^R(t) = e^{it}\epsilon_\tau(t),$$

since L just induces a rotation in both the q- and p-plane. There are now two cases

- ϵ_τ is a circle. In this case, ϵ_τ^R is periodic unless it is a critical point (which can happen if $\tau = 2\pi$).
- ϵ_τ is not a circle, in which case it is either a proper ellipse or a collision orbit (which looks like a line).

We now consider the second case, so the orbit ϵ_τ is not a circle. We then observe that such an orbit is periodic if the following resonance relation is satisfied for some positive integers k, ℓ

$$2\pi\ell = \tau k.$$

Hence periodic orbits in the rotating Kepler problem of the second kind have the following symmetry property.

Lemma 8.2.1. *Periodic orbits in the rotating Kepler problem of the second kind satisfy the following rotational symmetry,*

$$\epsilon_\tau^R(t+\tau) = e^{2\pi i\ell/k}\epsilon_\tau^R(t).$$

Proof. From the periodicity condition we have $\tau = 2\pi\ell/k$, so we find

$$\epsilon_\tau^R(t+\tau) = e^{it+i\tau}\epsilon_\tau(t+\tau) = e^{2\pi i\ell/k}e^{it}\epsilon_\tau(t) = e^{2\pi i\ell/k}\epsilon_\tau^R(t). \qquad \square$$

Figure 8.1 illustrates this lemma.

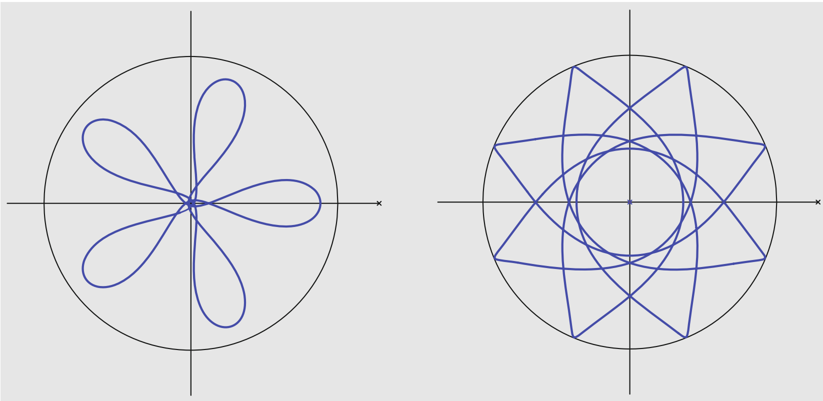


Figure 8.1: Periodic orbits in the rotating Kepler problem for $c = 1.6$: the circle indicates the boundary of the Hill's region.

Remark 8.2.2. It can be shown that periodic orbits of this type are dense in the set of bounded orbits in the rotating Kepler problem. One way to show this, is to use McGehee's disk-like surface of section, see McGehee's thesis [174] or the brief overview in Section 9.3, and write down the return map by integrating Equations (9.4). The resulting return map has the form

$$D^2 \longrightarrow D^2, \quad (r,\phi) \longmapsto (r, \phi + h_c(r)\,),$$

where h_c is an injective, smooth function. Whenever this function takes on a rational multiple of π as a value, we obtain a periodic orbit.

8.2.2 The circular orbits

If we fix H, the function

$$p_H(E) := p(H,E)$$

is a cubic polynomial in E and if we fix E, the function

$$p^E(H) := p(H,E)$$

is a quadratic polynomial in H. By Corollary 5.6.2 we know that $-\frac{3}{2}$ is the unique critical value of H. At the critical value the cubic polynomial $p_{-\frac{3}{2}}$ splits as follows

$$p_{-\frac{3}{2}}(E) = 2(E+2)\left(E+\tfrac{1}{2}\right)^2, \tag{8.3}$$

i.e., $p_{-\frac{3}{2}}$ has a simple zero at -2 and a double zero at $-\frac{1}{2}$. Recall that the discriminant of a cubic polynomial $p = ax^3 + bx^2 + cx + d$ is given by the formula

$$\Delta(p) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd.$$

In the case that the coefficients are real, the discriminant can be used to determine the number of real roots. Namely if $\Delta(p) > 0$ the polynomial p has three distinct real roots, if $\Delta(p) = 0$ the polynomial has a double root, and if $\Delta(p) < 0$ the polynomial has one real and two complex conjugated roots. For the cubic polynomial p_H the discriminant is computed as follows

$$\Delta(p_H) = 64H^6 - 64H^6 + 256H^3 - 108 - 288H^3 = -32H^3 - 108.$$

We see that $\Delta(p_H) > 0$ for $H < -\frac{3}{2}$, vanishes at $H = -\frac{3}{2}$ and satisfies $\Delta(p_H) < 0$ for $H > -\frac{3}{2}$. For $H < -\frac{3}{2}$ we denote by $E^1(H), E^2(H), E^3(H) \in \mathbb{R}$ the zeros of p_H ordered such that

$$E^1(H) < E^2(H) < E^3(H).$$

In view of (8.3) the three functions extend continuously to $H = -\frac{3}{2}$ such that

$$E^1\left(-\tfrac{3}{2}\right) = -2, \qquad E^2\left(-\tfrac{3}{2}\right) = E^3\left(-\tfrac{3}{2}\right) = -\tfrac{1}{2}.$$

Moreover, E^1 extends to a continuous function on the whole real line such that $E^1(H)$ is the unique real root of p_H if $H > -\frac{3}{2}$. Note that the discriminant of the quadratic polynomial $p^E(H) = 2EH^2 - 4E^2H + 2E^3 + 1$ equals

$$\Delta(p^E) = -8E$$

and therefore for $E < 0$ the polynomial p^E has precisely two real zeros. We conclude that the functions E^1 and E^2 are monotone increasing and the function E^3 is monotone decreasing such that

$$\lim_{H\to-\infty} E^1(H) = \lim_{H\to-\infty} E^2(H) = -\infty, \qquad \lim_{H\to-\infty} E^3(H) = \lim_{H\to\infty} E^1(H) = 0.$$

For later reference we observe that the images of the three functions are

$$\operatorname{im} E^1|_{(-\infty,\frac{3}{2}]} = (-\infty, 2], \quad \operatorname{im} E^2 = (-\infty, \tfrac{1}{2}], \quad \operatorname{im} E^3 = [\tfrac{1}{2}, 0). \tag{8.4}$$

Note that as an unparametrized, simple orbit a circular orbit in the (non-rotating) planar Kepler problem is uniquely determined by its energy E and its angular momentum L. On the other hand, for given values of E and L a circular periodic orbit only exists if $0 = 1 + 2EL^2$. That means that for a given negative energy value there exist precisely two circular orbits whose angular momenta differ by a sign, i.e., the circle is traversed backwards.

A circular periodic orbit of the Kepler problem also gives rise to a circular periodic orbit in the rotating Kepler problem, since the circle is invariant under rotation. In particular, a circular periodic orbit of the rotating Kepler problem is uniquely determined by the values of H and L or equivalently by the values of H and E.

Given a value of H and E, the existence of a circular periodic orbit implies that $p(H, E) = 0$ as defined in (8.1). Hence by the discussion above, for a given value of H less than the critical value $-\frac{3}{2}$ there exist three circular periodic orbits, while for a given energy value H bigger than the critical value $-\frac{3}{2}$ there exists a unique circular periodic orbit.

If the energy value c is less than $-\frac{3}{2}$ the Hill's region $\mathfrak{K}_c$ has two connected components, one bounded and one unbounded. We next discuss which of the three circular periodic orbits lie above the bounded component and which lie above the unbounded one. Because the Runge–Lenz vector for a circular periodic orbit vanishes, we obtain from (3.8) for the radius r of a circular periodic orbit

$$r = L^2 = -\frac{1}{2E}$$

while the second equality follows from (3.14) again in view of the fact that the Runge–Lenz vector vanishes. With Equation (8.4) we conclude that the circular periodic orbits corresponding to the energy values $E^1(c)$ and $E^2(c)$ have radius less than one, whereas the circular periodic orbit corresponding to the energy value $E^3(c)$ has radius bigger than one. Therefore the first two circular periodic orbits lie above the bounded component of the Hill's region where the third one lies above the unbounded component of the Hill's region.

The circular periodic orbit corresponding to E^1 is referred to as the *retrograde circular periodic orbit*, while the circular periodic orbit corresponding the E^2 is referred to as the (*interior*) *direct circular periodic orbit*. Finally the circular periodic orbit corresponding to E^3 is called the *exterior direct circular periodic orbit*. The exterior direct circular periodic orbit lies in the unbounded component while the other two orbits lie in the bounded component.

8.2.3 The averaging method

For very small energy the rotating Kepler problem approaches more and more the usual Kepler problem, which after Moser regularization is equivalent to the geodesic flow on the two-sphere. Due to invariance of the geodesic flow on the round two-sphere under rotation the closed geodesics are not isolated. We next explain how the circular retrograde and direct periodic orbits of the rotating Kepler problem bifurcate from the geodesic flow of the round two-sphere. This is known in the literature as the averaging method.

Periodic orbits can be interpreted variationally as critical points of Rabinowitz action functional. We first explain the bifurcation picture out of a Morse–Bott critical component in the finite-dimensional set-up. For this purpose suppose that X is a manifold and $f \in C^\infty(X \times [0,1), \mathbb{R})$. For $r \in [0,1)$ we abbreviate $f_r := f(\cdot, r) \in C^\infty(X, \mathbb{R})$ so that we obtain a one-parameter family of smooth functions on X. Suppose that

$$C \subset \operatorname{crit} f_0$$

is a Morse–Bott component of the critical set of f_0. We mean by this that $C \subset X$ is a closed submanifold which corresponds to a connected component of the critical set of f_0 with the property that for every $x \in C$ it holds that

$$T_x C = \ker H_{f_0}(x)$$

where $H_{f_0}(x)$ is the Hessian of f_0 at x. Suppose that the restriction of the derivative of f_r with respect to the homotopy variable r at $r = 0$ to the Morse–Bott component, which we will write as $\mathring{f}_0|_C$, is a Morse function. Then it follows from the implicit function theorem that there exists $\epsilon > 0$, an open neighborhood U of C in X and a smooth function

$$x \colon \operatorname{crit}(\mathring{f}_0|_C) \times [0, \epsilon) \to U$$

meeting the following conditions.

(i) If $\iota \colon \operatorname{crit}(\mathring{f}_0|_C) \to U$ is the inclusion and $x_0 = x(\cdot, 0) \colon \operatorname{crit}(\mathring{f}_0|_C) \to U$, then it holds that $x_0 = \iota$.

(ii) For every $r \in (0, \epsilon)$ the restriction $f_r|_U$ is Morse and we have $\operatorname{crit}(f_r|_U) = \operatorname{im} x_r$ where $x_r = x(\cdot, r) \colon \operatorname{crit}(\mathring{f}_0|_C) \to U$.

Recall from (5.22) that the Hamiltonian for the rotating Kepler problem reads

$$H \colon T^*(\mathbb{R}^2 \setminus \{0\}) \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|^2 - \frac{1}{|q|} + q_1 p_2 - q_2 p_1.$$

For an energy value $c < 0$ we regularize the rotating Kepler problem via

$$\begin{aligned} K^c(p,q) &:= \frac{1}{2}\Big(- \frac{|q|}{2c}\Big(H\big(\tfrac{q}{2c}, \sqrt{-2c}p\big) - c\Big) + 1\Big)^2 - \frac{1}{2} \\ &= \frac{1}{2}\Big(\frac{1}{2}(1 + |p|^2) + \frac{(q_1 p_2 - q_2 p_1)}{(-2c)^{\frac{3}{2}}}\Big)^2 |q|^2 - \frac{1}{2}. \end{aligned}$$

The discussion in Chapter 5.5.7 shows that the Hamiltonian K^c extends to a smooth Hamiltonian on T^*S^2 for every $c < 0$. By abuse of notation we denote the canonical smooth extension of K^c to T^*S^2 by the same letter. There is some small difference in the regularization above compared to the regularization in Chapter 4.4.1. In Chapter 4.4.1 we used the symplectic transformation $(p,q) \mapsto \big(-\frac{q}{\sqrt{-2c}}, \sqrt{-2c}p\big)$. Here we use the transformation $(p,q) \mapsto \big(\frac{q}{2c}, \sqrt{-2c}p\big)$ which is only conformally symplectic with conformal factor $\frac{1}{\sqrt{-2c}}$. For fixed c we can easily switch between the two transformations because a conformal symplectic factor can always be absorbed in the Hamiltonian in order to get the same Hamiltonian vector field. In particular, a conformal symplectic factor only gives rise to a reparametrization of the Hamiltonian flow. However, this transformation becomes problematic when one wants to study a sequence of symplectically conformal maps where the conformal factor converges to zero. This is precisely what we intend to do now. Namely we want to study the limit where c goes to $-\infty$. For that purpose we change the variable c in order to get the Hamiltonian

$$K_r(p,q) := K^{-\frac{1}{2r^{\frac{2}{3}}}}(p,q) = \frac{1}{2}\left(\frac{1}{2}(1+|p|^2) + (q_1p_2 - q_2p_1)r\right)^2 |q|^2 - \frac{1}{2}$$

for $r \in (0,\infty)$. This Hamiltonian smoothly extends to $r = 0$, where it becomes

$$K_0(p,q) = \frac{1}{2}\left(\frac{1}{2}(1+|p|^2)\right)^2 |q|^2 - \frac{1}{2}.$$

This is just the regularized Kepler Hamiltonian which coincides with kinetic energy on the round two-sphere. In particular, its Hamiltonian flow is just the geodesic flow on the round two-sphere.

We abbreviate by $\mathfrak{L} = C^\infty(S^1, T^*S^2)$ the free loop space of T^*S^2 and consider the family

$$\mathcal{A}_r := \mathcal{A}^{K_r} \colon \mathfrak{L} \times \mathbb{R}_+ \to \mathbb{R}$$

of action functionals as defined by (7.1). Critical points of $\mathcal{A}_0$ correspond to geodesics on the round two-sphere. Unparametrized simple closed geodesics on the round two-sphere are in one-to-one correspondence with unparametrized great circles. Great circles parametrized according to arc length are determined by a point and a unit direction. In particular, the space of parametrized great circles is diffeomorphic to $S^*S^2 \cong \mathbb{R}P^3$ where $S^*S^2 = \{v \in T^*S^2 : ||v|| = 1\}$ is the unit cotangent bundle of S^2. The circle S^1 acts on the space of parametrized great circles by time shift. Hence the space of unparametrized great circles is diffeomorphic to $\mathbb{R}P^3/S^1 \cong S^2$. If $\gamma\colon S^1 \to S^*S^2$ is a (parametrized) periodic orbit of the Hamiltonian vector field of K_0 corresponding to a simple closed geodesic, then because the period of a simple closed geodesic on the round two-sphere parametrized by arc length is 2π, the pair $(\gamma, 2\pi)$ is a critical point of $\mathcal{A}_0$. Abbreviate by

$$C \subset \mathrm{crit}\mathcal{A}_0$$

the space of all these pairs. Note that

$$C \cong \mathbb{R}P^3$$

is a Morse–Bott component of $\mathcal{A}_0$. The circle S^1 acts on $\mathfrak{L}$ by time-shift and on $\mathbb{R}_+$ trivially. Because the action functionals $\mathcal{A}_r$ are invariant under this S^1-action they induce action functionals

$$\overline{\mathcal{A}}_r\colon (\mathfrak{L} \times \mathbb{R}_+)/S^1 \to \mathbb{R}.$$

Note that S^1 acts on $\mathfrak{L} \times \mathbb{R}_+$ with finite isotropy so that the quotient $(\mathfrak{L} \times \mathbb{R}_+)/S^1$ is an orbifold. However, at $C \subset \mathfrak{L} \times \mathbb{R}_+$ the S^1-action is free and we denote the quotient by

$$\overline{C} = C/S^1 \subset \mathrm{crit}(\overline{\mathcal{A}}_0).$$

Note that

$$\overline{C} \cong \mathbb{R}P^3/S^1 \cong S^2.$$

We next study the restriction of $\overline{\mathring{\mathcal{A}}}_0$ to $\overline{C}$. Note that

$$\mathring{K}_0(p,q) = \frac{1}{2}(1+|p|^2)(q_1p_2 - q_2p_1)|q|^2 = \sqrt{2K_0+1}L|q| \tag{8.5}$$

where $L = q_1p_2 - q_2p_1$ is angular momentum. Note that at a point $(\gamma, \tau) \in \mathfrak{L} \times \mathbb{R}_+$ the derivative of $\mathcal{A}_r$ with respect to the r-variable at $r = 0$ is given by

$$\mathring{\mathcal{A}}_0(\gamma, \tau) = -\tau \int_{S^1} \mathring{K}_0(\gamma)dt.$$

If $(\gamma, 2\pi) = (p, q, 2\pi) \in C$ it follows that $K_0(\gamma) = 0$ and therefore

$$\mathring{\mathcal{A}}_0(\gamma, 2\pi) = -2\pi \int_{S^1} L(\gamma)|q|dt.$$

Because the angular momentum L is constant along periodic orbits of the Kepler problem, we can write this as

$$\mathring{\mathcal{A}}_0(\gamma, 2\pi) = -2\pi L(\gamma(0)) \int_{S^1} |q|dt.$$

The integral $\int_{S^1} |q|dt$ gives the ratio of the period of a Kepler ellipse of energy $-\frac{1}{2}$ before and after regularization. After regularization a Kepler ellipse becomes a closed geodesic on the round two-sphere and hence has period 2π. Before regularization by the version of the third Kepler law explained in Lemma 4.1.1 the period for energy $-\frac{1}{2}$ is 2π as well, so that $\int_{S^1} |q|dt = 1$ independent of the orbit. Therefore we can simplify the above formula to

$$\mathring{\mathcal{A}}_0(\gamma, 2\pi) = -2\pi L(\gamma(0)).$$

The induced map $\overline{L}$ of L on the quotient $S^*S^2/S^1 \cong \mathbb{R}P^3/S^1 \cong S^2$ is just the standard height function on the two-sphere. This is not a coincidence but a very special case of a much more general fact. Indeed, the Hamiltonian vector field of $\overline{L}$ on S^2 induces a periodic flow, so that we can think of $\overline{L}$ as a moment map for a circle action on S^2. By a very special case of the convexity theorem of Atiyah–Guillemin–Sternberg [21, 106] we know that such a moment map is Morse all whose critical points have even index. In particular, it has no saddle points and a unique maximum and a unique minimum. It is easy to see what the critical points are in our case. By the theory of Lagrange multipliers at a critical point of L on the constraint $K_0^{-1}(0)$ the differential of L and K_0 have to be proportional to each other, so that the Hamiltonian vector fields must be parallel. This happens precisely at the circular periodic orbits of the Kepler problem. For a fixed energy value there are precisely two circular periodic orbits moving in opposite direction.

8.2.4 Periodic orbits of the second kind

As we discussed in Section 8.2.1 apart from the circular orbits there are as well periodic orbits in the rotating Kepler problem which are rotating ellipses of positive eccentricity respectively rotating collision orbits. These periodic orbits are referred to as periodic orbits of the second kind. In order that a Kepler ellipse in the inertial system becomes a periodic orbit in the rotating or synodical system its period has to be a rational multiple of 2π, i.e.,

$$\tau = \frac{2\pi\ell}{k}$$

where $k, \ell \in \mathbb{N}$ are relatively prime. This is because the rotating coordinate system has period 2π with respect to the inertial system. The meaning of the positive integers k and ℓ is the following. While the coordinate system makes ℓ turns the ellipse makes k turns. These periodic orbits are never isolated. Indeed, we can rotate the whole periodic orbit to get a new periodic orbit. Therefore the periodic orbits of the second kind always appear in circle families. If we think of them as unparametrized simple orbits they appear actually in two-dimensional torus families. By Kepler's third law, namely Lemma 4.1.1 in Chapter 4, the energy $E_{k,\ell}$ of the ellipse is determined by its period through the formula

$$E_{k,\ell} = -\frac{1}{2}\left(\frac{k}{\ell}\right)^{\frac{2}{3}}.$$

If we fix the Jacobi energy H, then in view of $L = H - E$ the angular momentum is determined as well. If we know the energy and the angular momentum, then the eccentricity of the ellipse can be computed with the help of (3.14) as well. The shape of the ellipse up to rotation is then determined by (3.8) and because angular momentum cannot be zero for a proper ellipse the direction in which the ellipse is passed is then given, too. Summarizing we can say that if for given value c of H

and given relatively prime positive integers k and ℓ if there exists a corresponding periodic orbit of the second kind it is determined up to rotation. We refer to the pair (k, ℓ) as the *type* of the periodic orbit of the second kind. Astronomically the importance of periodic orbits of the second kind lies in the fact that in the Sun-Jupiter system asteroids often follow these orbits. That is the reason why for small integers k and ℓ these orbits actually have names which correspond to the names of representative asteroids lying on these orbits. Here are some of them.

Hecuba	Type $(2, 1)$.
Hilda	Type $(3, 2)$.
Thule	Type $(4, 3)$.
Hestia	Type $(3, 1)$.
Cybele	Type $(7, 4)$.

See Figure 8.2 for a plot of their shape. In general we call them just $T_{k,\ell}$. We next

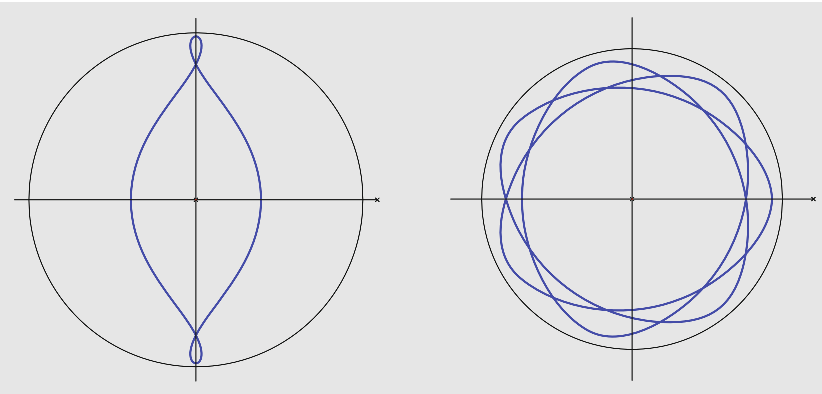


Figure 8.2: An older Hecuba ($c = 1.51$ on the left) and a young Cybele ($c = 1.55$ on the right).

determine the energy range for which these orbits exist. In view of (8.1) in order that a periodic orbit of type (k, ℓ) of energy c exists we need

$$(c - E_{k,\ell})^2 > -\frac{1}{2E_{k,\ell}}.$$

Abbreviating

$$L_{k,\ell} = \sqrt{-\frac{1}{2E_{k,\ell}}} = \left(\frac{\ell}{k}\right)^{\frac{1}{3}}$$

we set

$$c^-_{k,\ell} = E_{k,\ell} - L_{k,\ell} = -\frac{1}{2}\left(\frac{k}{\ell}\right)^{\frac{2}{3}} - \left(\frac{\ell}{k}\right)^{\frac{1}{3}} = -\left(\frac{\ell}{k}\right)^{\frac{1}{3}}\left(\frac{k+2\ell}{2\ell}\right)$$

and

$$c^+_{k,\ell} = E_{k,\ell} + L_{k,\ell} = -\frac{1}{2}\left(\frac{k}{\ell}\right)^{\frac{2}{3}} + \left(\frac{\ell}{k}\right)^{\frac{1}{3}} = \left(\frac{\ell}{k}\right)^{\frac{1}{3}}\left(\frac{k-2\ell}{2\ell}\right).$$

With this notation we see that in order for a period orbit of type (k,ℓ) of energy c to exist we need

$$c \in (c^-_{k,\ell}, c^+_{k,\ell}).$$

If one thinks of c as some kind of "life"-parameter of $T_{k,\ell}$, then $T_{k,\ell}$ is born at $c = c^-_{k,\ell}$ out of a $|k-\ell|$-fold covered circular periodic orbit. At its birth its angular momentum is $-L_{k,\ell} < 0$ and therefore the circular periodic orbit is direct. To decide if it is the interior or exterior direct orbit we distinguish the following cases.

$k > \ell$ In this case $|L_{k,\ell}| < 1$ and therefore the direct periodic orbit is interior.

$k = \ell = 1$ In this case $c^-_{1,1} = -\frac{3}{2}$ is the critical value of the rotating Kepler problem and the exterior and interior direct orbit both collapse to the critical circle.

$k < \ell$ In this case $|L_{k,\ell}| > 1$ and therefore the direct periodic orbit is exterior.

As c increases, the eccentricity of $T_{k,\ell}$ starts to increase until the middle of the life of $T_{k,\ell}$. Indeed, at $c = \frac{c^-_{k,\ell}+c^+_{k,\ell}}{2} = E_{k,\ell}$ the eccentricity becomes one in view of (3.14) and therefore $T_{k,\ell}$ becomes a collision orbit. One might interpret this event as kind of "midlife crisis" of $T_{k,\ell}$. After that the eccentricity of $T_{k,\ell}$ starts to decrease. Note that while the angular momentum for $T_{k,\ell}$ was negative in the first part of its life it becomes positive in the second part. One might think of this that $T_{k,\ell}$ changes its "political attitude" after its "midlife crisis". From a prograde attitude in the first part of its life it changes to a retrograde attitude. Finally $T_{k,\ell}$ dies at $c = c^+_{k,\ell}$ at the $k+\ell$-fold covered retrograde circular orbit.

8.3 The retrograde and direct periodic orbit

8.3.1 Low energies

We have seen how the retrograde and direct periodic orbit bifurcate from the geodesic flow in the rotating Kepler problem. The phenomenon described for the rotating Kepler problem is much more general as explained in the book by Moser and Siegel [188], see also the papers by Conley [57] and Kummer [155]. It continues to hold if one adds to the rotating Kepler problem some additional velocity independent forces.

Here is the set-up. Let $\Omega \subset \mathbb{R}^2$ be an open subset containing the origin and $V\colon \Omega \to \mathbb{R}$ be a smooth function with the property that the origin is a critical point of V and $\mu > 0$. We consider the Hamiltonian

$$H := H_{V,\mu}\colon T^*(\Omega \setminus \{0\}) \to \mathbb{R}, \quad (q,p) \mapsto \frac{1}{2}|p|^2 - \frac{\mu}{|q|} + q_1p_2 - q_2p_1 + V(q).$$

An example of a Hamiltonian of this form is the Hamiltonian H_m in (5.23) which is obtained from the Hamiltonian of the restricted three-body problem by shifting coordinates, or Hill's lunar Hamiltonian. For a given energy value $c < 0$ we regularize H by introducing the Hamiltonian

$$\begin{aligned} K^c(p,q) &= \frac{1}{2}\bigg(-\frac{|q|}{2c}\Big(H\big(\tfrac{q}{2c},\sqrt{-2c}p\big)-c-V(0)\Big)+\mu\bigg)^2-\frac{\mu^2}{2} \\ &= \frac{1}{2}\left(\frac{1}{2}(1+|p|^2)+\frac{(q_1p_2-q_2p_1)}{(-2c)^{\frac{3}{2}}}-\frac{\big(V\big(\frac{q}{-2c}\big)-V(0)\big)}{2c}\right)^2|q|^2-\frac{\mu^2}{2}. \end{aligned}$$

As in the rotating Kepler problem we change the energy parameter and introduce

$$\begin{aligned} K_r(p,q) &:= K^{\frac{-r^{-\frac{2}{3}}}{2}}(p,q) \\ &= \frac{1}{2}\left(\frac{1}{2}(1+|p|^2)+(q_1p_2-q_2p_1)r+\big(V(qr^{\frac{2}{3}})-V(0)\big)r^{\frac{2}{3}}\right)^2|q|^2-\frac{\mu^2}{2}. \end{aligned}$$

Note that

$$K_0(p,q)=\frac{1}{2}\left(\frac{1}{2}(1+|p|^2)\right)^2|q|^2-\frac{\mu^2}{2}$$

does not depend on V. In particular, its flow on the energy hypersurface $K_0^{-1}(0)$ coincides with the geodesic flow on the round two-sphere up to reparametrization. Abbreviate

$$G_r(p,q):=\frac{1}{2}(1+|p|^2)+(q_1p_2-q_2p_1)r+\big(V(qr^{\frac{2}{3}})-V(0)\big)r^{\frac{2}{3}}$$

so that we can write

$$K_r(p,q)=\frac{1}{2}G_r(p,q)^2|q|^2-\frac{\mu^2}{2}. \tag{8.6}$$

For the first derivative of G_r with respect to the homotopy parameter r we get

$$\frac{\partial G_r}{\partial r}(p,q)=(q_1p_2-q_2p_1)+\frac{2}{3}\big\langle\nabla V(qr^{\frac{2}{3}}),q\big\rangle r^{\frac{1}{3}}+\frac{2\big(V(qr^{\frac{2}{3}})-V(0)\big)}{3r^{\frac{1}{3}}}$$

In particular,

$$\frac{\partial G_r}{\partial r}\bigg|_{r=0}(p,q)=q_1p_2-q_2p_1=L(q,p).$$

For the second derivative of G_r we obtain

$$\begin{aligned} \frac{\partial^2 G_r}{\partial r^2}(p,q) &= \frac{4}{9}\big\langle H_V(qr^{\frac{2}{3}})q,q\big\rangle+\frac{2\big\langle\nabla V(qr^{\frac{2}{3}}),q\big\rangle}{9r^{\frac{2}{3}}}+\frac{4\big\langle\nabla V(qr^{\frac{2}{3}}),q\big\rangle}{9r^{\frac{2}{3}}} \\ &\quad -\frac{2\big(V(qr^{\frac{2}{3}})-V(0)\big)}{9r^{\frac{4}{3}}} \\ &= \frac{4}{9}\big\langle H_V(qr^{\frac{2}{3}})q,q\big\rangle+\frac{2\big\langle\nabla V(qr^{\frac{2}{3}}),q\big\rangle}{3r^{\frac{2}{3}}}-\frac{2\big(V(qr^{\frac{2}{3}})-V(0)\big)}{9r^{\frac{4}{3}}}, \end{aligned}$$

where we have abbreviated the Hessian of V by H_V. Because the origin is a critical point of V, we conclude

$$\begin{aligned}\frac{\partial^2 G_r}{\partial r^2}\bigg|_{r=0}(p,q) &= \frac{4}{9}\big\langle H_V(0)q,q\big\rangle + \frac{2}{3}\big\langle H_V(0)q,q\big\rangle - \frac{1}{9}\big\langle H_V(0)q,q\big\rangle \\ &= \big\langle H_V(0)q,q\big\rangle.\end{aligned}$$

In view of (8.6) we get for the first derivative of K_r

$$\frac{\partial K_r}{\partial r}(p,q) = G_r(p,q)|q|^2\frac{\partial G_r}{\partial r}(p,q) = \sqrt{(2K_r(p,q)+\mu^2)}|q|\frac{\partial G_r}{\partial r}(p,q)$$

so that we have

$$\frac{\partial K_r}{\partial r}\bigg|_{r=0}(p,q) = \sqrt{(2K_0+\mu^2)}|q|L.$$

In particular, this does not depend on V and therefore coincides with the computation for the rotating Kepler problem (8.5). For the second derivative of K_r it holds that

$$\begin{aligned}\frac{\partial^2 K_r}{\partial r^2} &= \frac{|q|}{\sqrt{2K_r+\mu^2}}\frac{\partial K_r}{\partial r}\frac{\partial G_r}{\partial r} + \sqrt{2K_r+\mu^2}|q|\frac{\partial^2 G_r}{\partial r^2} \\ &= |q|^2\left(\left(\frac{\partial G_r}{\partial r}\right)^2 + G_r\frac{\partial^2 G_r}{\partial r^2}\right).\end{aligned}$$

In particular, because G_r is two times continuously differentiable in r for every $r \in [0,\infty)$, the same is true for K_r.

As for the rotating Kepler problem we consider for $r \in [0,\infty)$ the family of action functionals

$$\mathcal{A}_r := \mathcal{A}^{K_r} : \mathfrak{L} \times \mathbb{R}_+ \to \mathbb{R}$$

as defined by (7.1) where $\mathfrak{L} = C^\infty(S^1, T^*S^2)$. Because K_r is twice differentiable and K_0 as well as $\mathring{K}_0$ do not depend on the choice of V, we conclude that the geodesic flow bifurcates at $r=0$ into two periodic orbits, precisely as in the rotating Kepler problem. We will still refer to these orbits as the direct and retrograde periodic orbits.

8.3.2 Birkhoff's shooting method

Recall from (5.27) that trajectories of Hill's lunar problem satisfy the following second-order ODE

$$\begin{cases} q_1'' + 2q_2' = q_1\left(3 - \frac{1}{|q|^3}\right) \\ q_2'' - 2q_1' = -\frac{q_2}{|q|^3} \end{cases} \tag{8.7}$$

The energy constraint for Hill's lunar problem becomes

$$c = \frac{1}{2}\big((q_1')^2 + (q_2')^2\big) - \frac{1}{|q|} - \frac{3}{2}q_1^2. \tag{8.8}$$

The following theorem is due to Birkhoff [39].

Theorem 8.3.1. *Assume $c < -\frac{3^{\frac{4}{3}}}{2}$. Then there exists $\tau > 0$ and $(q_1, q_2)\colon [0,\tau] \to (-\infty, 0] \times (-\infty, 0]$ solving* (8.7), (8.8), *and*

$$q_2(0) = 0, \quad q_1'(0) = 0, \quad q_1(\tau) = 0, \quad q_2'(\tau) = 0.$$

Proof. Consider the function

$$f_c\colon (0,\infty) \to \mathbb{R}, \quad r \mapsto c + \frac{1}{r} + \frac{3}{2}r^2.$$

Its derivative

$$f_c' = -\frac{1}{r^2} + 3r$$

has a unique zero at $r = \frac{1}{3^{\frac{1}{3}}} = 3^{-\frac{1}{3}}$ and satisfies

$$f_c'|_{\left(0,3^{-\frac{1}{3}}\right)} < 0, \quad f_c'|_{\left(3^{-\frac{1}{3}},\infty\right)} > 0.$$

In particular, f_c has a unique minimum at $r = 3^{-\frac{1}{3}}$ at which it attains the value

$$f_c(3^{-\frac{1}{3}}) = c + \frac{3^{\frac{4}{3}}}{2} < 0.$$

We conclude that there exists a unique $r_c \in (0, 3^{-\frac{1}{3}})$ such that

$$f_c(r_c) = 0.$$

Choose $r \in (0, r_c]$. Let $q^r\colon [0, T_r) \to \mathbb{R}^2$ be the solution of (8.7) to the initial conditions

$$q_1^r(0) = -r, \quad q_2^r(0) = 0, \quad (q_1^r)'(0) = 0, \quad (q_2^r)'(0) = -\sqrt{2f_c(r)}. \tag{8.9}$$

In view of the initial conditions (8.8) holds for every t by preservation of energy. In particular, q^r lies in the bounded part of the Hill's region whose only non-compactness comes from collisions at the origin. Hence we choose $T_r \in (0, \infty]$ such that $\lim_{t\to T_r} q^r(t) = 0$ in case that T_r is finite. We introduce the quantity

$$\tau(r) := \inf\big\{t \in (0, T_r) : q_2^r(t) = 0, \text{ or } q_1^r(t) = 0\big\}.$$

Here we understand that if the set is empty, then $\tau(r) = T_r$. If $r < r_c$, then in view of the initial conditions (8.9) we have $(q_2^r)'(0) < 0$ and therefore

$$\tau(r) > 0.$$

We claim that

$$\tau(r) < \infty, \quad r \in (0, r_c). \tag{8.10}$$

To see that, we first integrate the first equation in (8.7) and use the initial condition (8.9)

$$\begin{aligned}(q_1^r)'(t) &= (q_1^r)'(0) - 2q_2^r(t) + 2q_2^r(0) + \int_0^t q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)ds \\ &= -2q_2^r(t) + \int_0^t q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)ds.\end{aligned} \tag{8.11}$$

We further note that by the initial condition and the definition of $\tau(r)$ it holds that

$$q_2^r(t) < 0, \quad q_1^r(t) < 0, \qquad 0 < t < \tau(r). \tag{8.12}$$

If $\mathfrak{K}_c^b$ is the bounded part of the Hill's region we claim further that

$$\mathfrak{K}_c^b \subset B_{3^{-\frac{1}{3}}}(0), \tag{8.13}$$

i.e., the bounded part of the Hill's region is contained in the ball of radius $3^{-\frac{1}{3}}$ centered at the origin. To prove (8.13) suppose that

$$(q_1, q_2) \in \partial B_{3^{-\frac{1}{3}}}(0).$$

In particular,

$$|q| = \frac{1}{3^{\frac{1}{3}}}.$$

We estimate

$$-\frac{1}{|q|} - \frac{3}{2}q_1^2 \geq -\frac{1}{|q|} - \frac{3}{2}|q|^2 = -\frac{3^{\frac{4}{3}}}{2} > c.$$

In view of the characterization (5.30) of the Hill's region as a sublevel set, this implies that

$$\partial B_{3^{-\frac{1}{3}}}(0) \cap \mathfrak{K}_c^b = \emptyset.$$

Because $\mathfrak{K}_c^b$ is connected and contains the origin in its closure, the inclusion (8.13) follows. Combining (8.13) with the second inequality in (8.12) we conclude that

$$q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)(t) > 0, \quad 0 < t < \tau(r). \tag{8.14}$$

Because $q_1^r(0) = -r$ there exists $t_0 > 0$ such that

$$q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)(t) \geq \mu > 0, \quad 0 \leq t \leq t_0. \tag{8.15}$$

For $t_0 \leq t < \tau(r)$ we conclude from (8.11), (8.14), and (8.15) in combination with the first inequality in (8.12) that

$$(q_1^r)'(t) = -2q_2^r(t) + \int_0^t q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)ds \geq \int_0^{t_0} q_1^r\Big(3 - \tfrac{1}{|q^r|^3}\Big)ds \geq \mu t_0. \tag{8.16}$$

This implies that

$$q_1^r(t) \geq q_1^r(t_0) + \mu t_0(t - t_0), \quad t_0 \leq t < \tau(r).$$

Because the Hill region $\mathfrak{K}_c^b$ is bounded, $\tau(r)$ is finite and (8.10) is proved.

We define

$$\rho := \inf\left\{r \in (0, r_c) : q_1^r(\tau(r)) = 0,\ q_2^r(\tau(r)) = 0\right\}$$

with the convention that $\rho = r_c$ if the set is empty. We claim that

$$q_1^r(\tau(r)) = 0, \quad r \in (0, \rho). \tag{8.17}$$

In order to prove (8.17) we introduce the quantity

$$r_0 := \inf\left\{r \in (0, \rho) : q_2^r(\tau(r)) = 0\right\}$$

with the convention that $r_0 = \rho$ if the set is empty. We need to show that $r_0 = \rho$. For small r the trajectory q_1^r is close to the origin and we conclude from the dynamics of the Kepler problem that $q_1^r(\tau(r)) = 0$. In particular,

$$r_0 > 0.$$

We now argue by contradiction and assume that $r_0 < \rho$. That means that

$$q_2^{r_0}(\tau(r_0)) = 0, \quad (q_2^{r_0})'(\tau(r_0)) = 0, \quad (q_2^{r_0})''(\tau(r_0)) \leq 0.$$

By (8.16) we have

$$(q_1^{r_0})'(\tau(r_0)) > 0.$$

This contradicts the second equation in (8.7) and (8.17) is proved.

Our next claim is

$$\rho < r_c. \tag{8.18}$$

To prove that, we consider the trajectory q^{r_c}. Its initial conditions are

$$q_1^{r_c}(0) = -r_c, \quad q_2^{r_c}(0) = 0, \quad (q_1^{r_c})'(0) = 0, \quad (q_2^{r_c})'(0) = 0.$$

The second equation in (8.7) implies that

$$(q_2^{r_c})''(0) = 2(q_1^{r_c})'(0) - \frac{q_2^{r_c}(0)}{|q^{r_c}(0)|^3} = 0.$$

Differentiating the second equation in (8.7) and using $(q_2^{r_c})'(0) = (q_1^{r_c})'(0) = 0$ we conclude

$$\begin{aligned}
(q_2^{r_c})'''(0) &= 2(q_1^{r_c})''(0) \\
&= -4(q_2^{r_c})'(0) + 2q_1^{r_c}(0)\left(3 - \frac{1}{|q^{r_c}(0)|^3}\right) \\
&= -2r_c\left(3 - \frac{1}{r_c^3}\right) \\
&> 0.
\end{aligned}$$

Here we have used for the second equality the second equation in (8.7) and for the last inequality the fact that $r_c \in (0, 3^{-\frac{1}{3}})$. Summarizing we have

$$q_2^{r_c}(0) = 0, \quad (q_2^{r_c})'(0) = 0, \quad (q_2^{r_c})''(0) = 0, \quad (q_2^{r_c})'''(0) > 0.$$

In particular, there exists $\epsilon > 0$ such that

$$q_2^{r_c}(t) > 0, \quad t \in (0, \epsilon). \tag{8.19}$$

We assume now by contradiction that $\rho = r_c$. It follows from (8.17) that

$$q_1^r(\tau(r)) = 0, \quad r \in (0, r_c).$$

This implies that the above $\epsilon > 0$ can be chosen so small such that

$$\tau(r) \geq \epsilon, \quad r \in [\tfrac{r_c}{2}, r_c).$$

In particular,

$$q_2^r(t) \leq 0, \quad r \in [\tfrac{r_c}{2}, r_c), \; t \in (0, \epsilon).$$

But this has the consequence that

$$q_2^{r_c}(t) \leq 0, \quad t \in (0, \epsilon)$$

in contradiction to (8.19). This contradiction proves (8.18).

Together with (8.18), the dynamics of the Kepler problem tell us that, for r close to ρ, we have $(q_2^r)'(\tau(r)) > 0$. On the other hand the dynamics of the Kepler problem also implies that $(q_2^r)'(\tau(r)) < 0$ for r close to 0. By the intermediate value theorem we conclude that there exists $r \in (0, \rho)$ such that

$$(q_2^r)'(\tau(r)) = 0.$$

This proves the theorem. □

With more effort such a shooting argument can also be made to work for the restricted three-body problem. The upshot, also due to Birkhoff, is the existence of a retrograde orbit for all mass ratios $\mu < 1$ for energies below the first critical point. In his proof in [39], Birkhoff uses for the restricted three-body problem actually a double shooting argument. Because the restricted three-body problem is not invariant anymore under reflection at the q_2-axis, Birkhoff shoots perpendicularly at the q_1-axis from the right and left of the mass and shows that one can arrange the initial conditions such that the two shots meet. A different argument can be found in the paper by Conley [58]. Conley applies the Levi-Civita regularization to the restricted three-body problem. After Levi-Civita regularization the problem is again invariant under reflection at the q_1- and q_2-axis like Hill's lunar problem and as in Hill's lunar problem one needs to show that one can shoot perpendicularly from the q_1-axis to hit the q_2-axis perpendicularly.

To state the result, we first abbreviate for $\mu \in (0,1)$

$$\kappa_\mu := H_\mu(L_1^\mu)$$

the first critical value of the Hamiltonian H_μ of the restricted three-body problem for mass ratio μ, namely its value at the first Lagrange point L_1^μ, see Theorem 5.4.7 in Chapter 5. Note that $\kappa_\mu = \kappa_{1-\mu}$ and κ_μ smoothly extends to $\kappa_0 = \kappa_1 = -\frac{3}{2}$, the critical value of the rotating Kepler problem. We further recall that the critical value can also be obtained as $\kappa_\mu = U_\mu(\ell_1^\mu)$, where U_μ is the effective potential and ℓ_1^μ is the critical point of U_μ obtained by projecting L_1^μ to position space. Birkhoff's theorem can now be stated as follows.

Theorem 8.3.2 (Birkhoff)**.** *Assume that $\mu \in (0,1)$ and $c \in (-\infty, \kappa_\mu)$, then there exists $\tau > 0$ and $(q_1, q_2)\colon [0,\tau] \to \mathbb{R} \times (-\infty, 0]$ solving* (5.7)*, namely the second-order ODE of trajectories of the restricted three-body problem for mass ratio μ, the energy constraint*

$$c = \frac{1}{2}|q'|^2 + U_\mu(q)$$

as well as the boundary conditions

$$q_2(0) = q_2(\tau) = 0, \quad q_1'(0) = q_1'(\tau) = 0, \quad \ell_3^\mu < q_1(0) < -\mu, \quad -\mu < q_1(\tau) < \ell_1^\mu.$$

Although Birkhoff uses in his proof on the existence of the retrograde periodic orbit in the restricted three-body problem a double shooting argument a single shooting procedure as in Hill's lunar problem can be implemented on a computer. See Figure 8.3. In fact, although we are not aware of an analytic proof of their existence Birkhoff's shooting method can also be applied to find direct orbits. In this case one does not shoot to $\mathbb{R} \times (-\infty, 0]$ but to $\mathbb{R} \times [0, \infty)$.

Remark 8.3.3. The boundary conditions tell us that q starts at the axis containing the earth (E) and the moon (M) from the left to the earth and ends on this axis to the right of the earth. That the starting point is right to the third Lagrange point ℓ_3 guarantees that q lies in the bounded component of the Hill's region. It is then clear that its endpoint has to lie to the left of the first Lagrange point ℓ_1. See Figure 8.3 for an illustration.

Numerical experiments by the second named author suggest that the solution guaranteed by Birkhoff is actually unique. Nevertheless we are not aware of an analytic proof of this fact. So we put it as an open question.

Question 8.3.4. *Is the solution obtained in Theorem* 8.3.2 *unique?*

To make the solution unique we pick the solution whose starting point is closest to the earth among all solutions. If we look at the corresponding solution in phase space after reparametrization, we have a brake orbit which we double to a periodic orbit. After Moser regularization we obtain a periodic orbit in a fiberwise starshaped hypersurface over S^2 which is diffeomorphic to $\mathbb{R}P^3$. Because

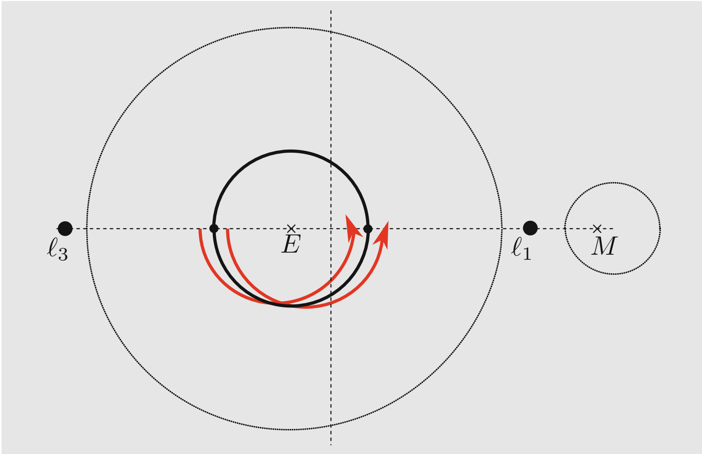


Figure 8.3: In the restricted three-body problem Birkhoff could still find a retrograde periodic orbit analytically. A cartoon of the argument is easy: we can still find a sign change in the slope, and use the symmetry to produce the black periodic orbit.

the starting and endpoint of q lie on different sides of the earth this periodic orbit is non-contractible. Hence only its double cover lifts to a periodic orbit on S^3. We denote this double cover by

$$\gamma_R = \gamma_R^{\mu,c} \colon [0,1] \to \Sigma_{\mu,c}$$

where $\Sigma_{\mu,c} \subset \mathbb{C}^2$ is the starshaped hypersurface we obtain by Levi-Civita regularization of the bounded component around the earth of the restricted three-body problem for mass ratio μ and energy c. Here to actually get a parametrized periodic orbit we can fix the parametrization by insisting that $\gamma_R(0) \in \mathbb{R} \times i\mathbb{R}$ and of course we parametrize it in such a way that γ_R becomes a minimal parametrized periodic orbit. By its construction it is symmetric with respect to the anti-symplectic involution $(z_1, z_2) \mapsto (\overline{z}_1, -\overline{z}_2)$ on $\mathbb{C}^2$ and because it is the double cover of a periodic orbit on $\mathbb{R}P^3$ it satisfies $\gamma_R\big(t + \frac{1}{2}\big) = -\gamma_R(t)$. In the following we refer to this orbit as the *retrograde periodic orbit*. For small energies this notion coincides with the one from Section 8.3. The following question is related but logically independent from Question 8.3.4.

Question 8.3.5. *Is the retrograde periodic orbit non-degenerate?*

Remark 8.3.6. In the question we mean non-degeneracy of the retrograde periodic orbit as a parametrized periodic orbit. There are weaker versions of non-degeneracy one could ask for. Because the retrograde periodic orbit is symmetric one could ask if the underlying brake orbit is non-degenerate. Indeed, an affirmative answer to the question would imply this as well by Lemma 7.5.5. Because the retrograde periodic orbit is a double cover of a non-contractible periodic orbit on the quotient diffeomorphic to $\mathbb{R}P^3$ one could ask for non-degeneracy of this orbit as well. Again an affirmative answer to the question implies this.

8.3.3 The Birkhoff set

We denote by

$$\mathfrak{C} = \big\{(\mu, c) \in (0,1) \times \mathbb{R} : c < \kappa_\mu\big\}$$

the two parameter space of mass ratios μ and energies c for which Theorem 8.3.2 guarantees the existence of the retrograde periodic orbit. We abbreviate

$$\mathfrak{B}_1 = \big\{(\mu, c) \in \mathfrak{C} : \gamma_R^{\mu,c} \text{ non-degenerate}\big\}.$$

Note that $\mathfrak{B}_1$ is an open subset of $\mathfrak{C}$. Moreover, by the results from Section 8.3 the retrograde periodic orbit is non-degenerate for very small energies and therefore there exists $c_1 \in \mathbb{R}$ such that

$$(0,1) \times (-\infty, c_1) \in \mathfrak{B}_1. \tag{8.20}$$

Moreover, as μ goes to zero, i.e., as we approach the rotating Kepler problem, the retrograde periodic orbit converges to the double cover of the retrograde periodic orbit of the rotating Kepler problem which is non-degenerate and therefore for each $c < -\frac{3}{2} = \kappa_0$ there exists $\mu_1(c) > 0$ such that

$$\Big\{(\mu, c) \in (0,1) \times \big(-\infty, -\tfrac{3}{2}\big) : \mu < \mu_1(c)\Big\} \subset \mathfrak{B}_1. \tag{8.21}$$

Numerical experiments of the second named author actually indicate that it might hold that $\mathfrak{B}_1 = \mathfrak{C}$. A nice property of the set $\mathfrak{B}_1$ is that if we have a smooth $\psi \colon (0,1) \to \mathfrak{B}_1$, then we obtain a smooth path of parametrized periodic orbits $r \mapsto \gamma_R^{\psi(r)}$ for $r \in (0,1)$.

We next introduce a further subset of the set $\mathfrak{B}_1$. For $(\mu, c) \in \mathfrak{C}$ abbreviate

$$\tau_R^{\mu,c} := \tau(\gamma_R^{\mu,c})$$

the period of the retrograde periodic orbit. Note that $\tau_R^{\mu,c}$ does not need to coincide with the period τ appearing in Theorem 8.3.2 because in the regularization process the orbit becomes reparametrized. Now abbreviate by

$$\mathfrak{B}_2 \subset \mathfrak{B}_1$$

the subset of all $(\mu, c) \in \mathfrak{B}_1$ which have the property that the retrograde periodic orbit $\gamma_R^{\mu,c}$ is linked to every periodic Reeb orbit on $\Sigma_{\mu,c}$ of period less than $\tau_R^{\mu,c}$. Observe that $\mathfrak{B}_2$ is an open subset of $\mathfrak{B}_1$ and therefore of $\mathfrak{C}$. Note further that a point $(\mu, c) \in \mathfrak{B}_1$ belongs to $\mathfrak{B}_2$ if the retrograde periodic orbit $\gamma_R^{\mu,c}$ has minimal period among all periodic Reeb orbits on $\Sigma_{\mu,c}$. In particular, this holds true for very low energies and for the rotating Kepler problem so that maybe after adjusting the constants we get analogous results for (8.20) and (8.21) namely there exists $c_2 \in \mathbb{R}$ with the property that

$$(0,1) \times (-\infty, c_2) \in \mathfrak{B}_2 \tag{8.22}$$

and $\mu_2(c) > 0$ for every $c \leq -\frac{3}{2}$ such that

$$\Big\{(\mu, c) \in (0,1) \times \big(-\infty, -\tfrac{3}{2}\big) : \mu < \mu_2(c)\Big\} \subset \mathfrak{B}_2. \tag{8.23}$$

In view of (8.22) we define the *Birkhoff set*

$$\mathfrak{B} \subset \mathfrak{B}_2$$

as the connected component of $\mathfrak{B}_2$ having the property that

$$(0,1) \times (-\infty, c_2) \in \mathfrak{B}.$$

From (8.23) we deduce that

$$\Big\{(\mu, c) \in (0,1) \times \big(-\infty, -\tfrac{3}{2}\big) : \mu < \mu_2(c)\Big\} \subset \mathfrak{B}.$$

Note that $\mathfrak{B}$ is a connected open subset of $\mathfrak{C}$ and therefore in particular, it is path-connected. The importance of the Birkhoff set is that, as we will show, for every $(\mu, c) \in \mathfrak{B}$ the energy hypersurface $\Sigma_{\mu,c} \subset \mathbb{C}^2$ admits a global surface of section. The question we have is

Question 8.3.7. *Is $\mathfrak{B} = \mathfrak{C}$?*

Of course Question 8.3.7 is equivalent to the question if $\mathfrak{B}_2 = \mathfrak{C}$, i.e, it contains Question 8.3.5 together with the question if all orbits of period less than the period of the retrograde periodic orbit are linked to it. An affirmative answer to the second question would be implied by an affirmative answer to the following question

Question 8.3.8. *Does the retrograde periodic orbit have minimal period among all periodic Reeb orbits on $\Sigma_{\mu,c}$?*

A starshaped hypersurface in $\mathbb{C}^2$ is called dynamically convex if the Conley–Zehnder indices of all periodic orbits are bigger or equal to three. We say that $\Sigma_{\mu,c}$ is *weakly dynamically convex* if all periodic orbits of period less than $\tau_R^{\mu,c}$ have Conley–Zehnder index at least three. We abbreviate

$$\mathfrak{A}_1 := \big\{(\mu, c) \in \mathfrak{B}_1 : \Sigma_{\mu,c} \text{ weakly dynamically convex}\big\}.$$

In fact, for small energy values $\Sigma_{\mu,c}$ is dynamically convex and we abbreviate by

$$\mathfrak{A} \subset \mathfrak{A}_1$$

the connected component of $\mathfrak{A}_1$ containing the small energy values. We prove

Theorem 8.3.9. $\mathfrak{A} \subset \mathfrak{B}$.

In particular, this prompts the question

Question 8.3.10. *Is $\Sigma_{\mu,c}$ dynamically convex for every $(\mu, c) \in \mathfrak{C}$?*

Note that in view of Theorem 8.3.9 an affirmative answer to Question 8.3.5 and Question 8.3.10 would provide an affirmative answer to Question 8.3.7. Indeed, suppose that Question 8.3.5 is answered affirmatively. Then it follows that $\mathfrak{B}_1 = \mathfrak{C}$. If Question 8.3.10 is answered affirmatively as well, we obtain $\mathfrak{A}_1 = \mathfrak{C}$. But $\mathfrak{C}$ is connected so that we have $\mathfrak{A} = \mathfrak{A}_1 = \mathfrak{C}$. From Theorem 8.3.9 we deduce that $\mathfrak{B} = \mathfrak{C}$.

8.4 Periodic orbits of the second kind for small mass ratios

We have seen that in the rotating Kepler problem there exist two kinds of periodic orbits, the circular periodic orbits and the elliptic ones. The first type is referred to by Poincaré [205] as periodic orbits of the first kind and the second type as periodic orbits of the second kind. We discuss how they survive for small mass ratios. In order to do so we have to check whether the periodic orbits in the rotating Kepler problem are non-degenerate. Because the rotating Kepler problem is completely integrable, periodic orbits are usually non-isolated and non-degeneracy only has a chance to hold if one takes the symmetry into account as in Definition 7.3.3 in Chapter 7. To visualize the symmetry we use the Ligon–Schaaf embedding of the Kepler problem we discussed in Section 4.3 in Chapter 4. In the following we think of the two-dimensional sphere as embedded into the three-dimensional Euclidean space $\mathbb{R}^3$ as the subset $S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$. Consider the map

$$\nu \colon T^*S^2 \setminus S^2 \to \mathbb{R}^2, \quad (x, y) \mapsto \big(2\pi|y|, 2\pi(x_1y_2 - x_2y_1)\big).$$

The flow of the Hamiltonian vector field of the first component ν_1 of ν is the geodesic flow on the round two-dimensional sphere which is periodic of period one and the flow of the Hamiltonian vector field of the second component ν_2 is the physical symmetry obtained by rotation of S^2 along the third axis which is periodic of period one as well. Moreover, the two flows commute so that we can think of ν as the moment map of a torus action on $T^*S^2 \setminus S^2$. In view of property (ii) of the Ligon–Schaaf embedding Φ discussed in Section 4.3 in Chapter 4 the Hamiltonian E of the (non-rotating) Kepler problem satisfies

$$E = -\frac{2\pi^2}{\nu_1^2 \circ \Phi}.$$

By the third property of the Ligon–Schaaf map it further holds that the angular momentum satisfies

$$L = \frac{\nu_2 \circ \Phi}{2\pi}.$$

Let $f\colon \operatorname{im}(\nu) \to \mathbb{R}$ be the function

$$f(\nu_1, \nu_2) = \frac{2\pi^2}{\nu_1^2} + \frac{\nu_2}{2\pi}$$

and abbreviate

$$\nu_f := f \circ \nu\colon T^*S^2 \setminus S^2 \to \mathbb{R}.$$

Then in view of the discussion above the Hamiltonian of the rotating Kepler problem satisfies

$$H = E + L = \nu_f \circ \Phi.$$

Note that periodic orbits of the second kind as discussed in Section 8.2.4 in the rotating Kepler problem, namely rotating ellipses of positive eccentricity together with rotating collision orbits, are identified via the Ligon–Schaaf map with periodic orbits of the second kind on the T^2-manifolds $T^*S^2 \setminus S^2$ as discussed in Definition 7.4.2 in Chapter 7, namely periodic orbits in the regular set of the moment map ν. In view of Example 7.4.6 in Chapter 7 the function f satisfies hypothesis $(H2)$ so that by Theorem 7.4.7 in the same chapter each periodic orbit of the second kind of ν_f is non-degenerate, when interpreted as a periodic orbit on the T^2-manifold $T^*S^2 \setminus S^2$.

Recall that the Hamiltonian of the restricted three-body problem is invariant under the anti-symplectic involution

$$\rho_0\colon T^*\mathbb{R}^2 \to T^*\mathbb{R}^2, \quad (q_1, q_2, p_1, p_2) \mapsto (q_1, -q_2, -p_1, p_2).$$

Like in Section 6.3 in Chapter 6 consider the anti-symplectic involution

$$\rho\colon T^*S^2 \to T^*S^2$$

obtained by the composition of d_*I, where I is the reflection at the great arc $\{x \in S^2 : x_1 = 0\}$, and the anti-symplectic involution $(x, y) \mapsto (x, -y)$. Explicitly, ρ is given by the formula

$$\rho(x_1, x_2, x_3, y_1, y_2, y_3) = (-x_1, x_2, x_3, y_1, -y_2, -y_3)$$

for $(x, y) \in T^*S^2 \subset T^*\mathbb{R}^3 = \mathbb{R}^3 \times \mathbb{R}^3$. Like the Moser regularization, the Ligon–Schaaf regularization interchanges the anti-symplectic involutions ρ_0 and ρ, i.e.,

$$\rho \circ \Phi = \rho_0. \tag{8.24}$$

Because the Moser regularization and the Ligon–Schaaf regularization are different regularizations this might look surprising. However, (8.24) can be readily checked using the explicit description of the Ligon–Schaaf map in Section 4.3 in Chapter 4 or alternatively the description of the Ligon–Schaaf map using Moser regularization in [112] can be used. Note that

$$\nu \circ \rho = \nu$$

so that ν becomes a real moment map on the real symplectic manifold $(T^*S^2 \setminus S^2, \rho)$.

The elliptic periodic orbits of the rotating Kepler problem arise in whole circle families, namely one can rotate the ellipse. Using the rotation we can find a symmetric representative. As we noted above periodic orbits of the second kind are non-degenerate as periodic orbits on the T^2-manifold $T^*S^2 \setminus S^2$. Hence in view of Lemma 7.5.7 in Chapter 7 its symmetric representatives are non-degenerate as symmetric periodic orbits. In particular, they survive for small values of μ. The idea of this proof of existence of periodic orbits of the second kind is essentially due to Poincaré [205]. Rigorous arguments were provided by Arenstorf and Barrar, see [15, 25, 26]. To relate the existence of periodic orbits of the second kind directly to the geometry of the moment map via the Ligon–Schaaf regularization seems to be new. Nevertheless, in disguise these arguments are already contained in the classical approaches. For example, the argument by Barrar [25] uses Delaunay coordinates which are a kind of local version of the Ligon–Schaaf regularization. For different arguments concerning periodic orbits of the second kind we recommend to have a look at the excellent treatments in [57], [110, Chapter 6] or [189, Chapter 2].

In view of the discussion above there are exist many symmetric periodic orbits in the restricted three-body problem for small mass ratios μ, which converge as μ goes to zero to our beloved friends from the rotating Kepler problem like Hecuba, Hilda, Thule or Hestia. Given a simple symmetric periodic orbit in the restricted three-body problem below the first critical value in the bounded component, say around the earth, we would like to know if the periodic orbit is contractible or not in the regularized hypersurface which is diffeomorphic to $\mathbb{R}P^3$, which has fundamental group $\mathbb{Z}/2\mathbb{Z}$. Suppose that our periodic orbit is not a collision, so that we can discuss it already in the unregularized problem. By Lemma 7.5.2 in Chapter 7 we know that a symmetric periodic orbit intersects the fixed point set of the anti-symplectic involution in precisely two points. In the unregularized problem this means that the projection of a periodic orbit to position space has precisely two perpendicular intersections with the axis of syzygies, namely the axis containing earth and moon. Now there are two possibilities. Namely the two perpendicular intersections occur at the same side of the earth or at two different sides. By Proposition 7.5.8 in Chapter 7 in the first case the orbits is contractible while in the second one it is not contractible.

8.5 Lyapunov orbits

Theorem 8.5.1. *For any $\mu \in (0,1)$ there exists $\epsilon = \epsilon(\mu) > 0$ such that for every collinear Lagrange point L_i for $i \in \{1,2,3\}$ there exists a smooth family of periodic orbits $\gamma_i^c \colon S^1 \to H^{-1}(c)$ where $c \in \big(H(L_i), H(L_i) + \epsilon\big)$ and H is the Hamiltonian of the restricted three-body problem such that $\gamma_i^c(t)$ converges uniformly to L_i as c goes to $H(L_i)$.*

Remark 8.5.2. The periodic orbits γ_i^c are known as *Lyapunov orbits*.

The question under which conditions there exist Lyapunov orbits at the equilateral Lagrange points L_4 and L_5 is trickier, see for example [110, Chapter 6]. The reason for this is that L_1, L_2, and L_3 have Morse index 1, where L_4 and L_5 have Morse index 2. This holds because the projections to position space $\ell_i = \pi(L_i)$ for $i \in \{1, 2, 3\}$ are saddle points of the effective potential U while ℓ_4 and ℓ_5 are maxima as we discussed in Theorem 5.4.7 in Chapter 5.

The Lagrange points are critical points of the Hamiltonian H and therefore also singularities of the Hamiltonian vector field X_H. By linearizing at these points, we end up with some fascinating linear algebra. We first discuss a more general set-up. Suppose that $\pi\colon E \to M$ is a vector bundle E over a manifold M and assume further that $s\colon M \to E$ is a section, i.e., $\pi \circ s = \mathrm{id}|_M$. Suppose that $x \in M$ is a zero of the section s, i.e., $s(x) = x$ if M is identified with the zero section in E and therefore becomes a subset $M \subset E$. The differential of s at x is a linear map

$$ds(x)\colon T_xM \to T_xE.$$

Because x lies in the zero section of E we have a canonical splitting

$$T_xE = T_xM \oplus E_x.$$

Abbreviate by

$$\Pi_x\colon T_xE \to E_x$$

the projection along T_xM. The *vertical differential* of s at x is defined to be

$$Ds(x) := \Pi_x \circ ds(x)\colon T_xM \to E_x.$$

This is nothing frightening at all. Just look at the trivial case $E = M \times \mathbb{R}^m$ where m is the rank of E. In this case we can write

$$s = \mathrm{id}|_M \oplus \sigma$$

where

$$\sigma\colon M \to \mathbb{R}^m$$

is a smooth map. Then we have $\sigma(x) = 0$ and

$$Ds(x) = d\sigma(x).$$

The differential of a smooth function $f\colon M \to \mathbb{R}$ can be thought of as a section

$$df\colon M \to T^*M.$$

A critical point x of f is then a zero of this section and the vertical differential is a linear map

$$Ddf(x)\colon T_xM \to T_x^*M.$$

This gives rise to a bilinear form

$$H_f(x)\colon T_xM \times T_xM \to \mathbb{R}$$

given for tangent vectors $v, w \in T_xM$ by

$$H_f(x)(v,w) = \big\langle Ddf(x)v, w\big\rangle.$$

This bilinear form is known as the *Hessian* of f at x and it turns out that it is symmetric. There are not many invariants of a symmetric bilinear form by the theorem of Sylvester, namely just the dimension of its kernel and its Morse index, namely the number of negative eigenvalues counted with multiplicity. Now consider a vector field V on M which can be thought of as a section

$$V\colon M \to TM.$$

If $x \in M$ is a zero of the vector field V, then the vertical differential becomes a linear map

$$DV(x)\colon T_xM \to T_xM.$$

Note that in this case we have many more invariants than in the case of the Hessian. In particular, the eigenvalues of $DV(x)$ are invariants.

At a Lagrange point L_i for $i \in \{1, \dots, 5\}$ we obtain from the definition of the Hamiltonian vector field the relation

$$DdH(L_i) = \omega(\cdot, DX_H(L_i)\cdot). \tag{8.25}$$

The proof of Theorem 8.5.1 hinges on the fact that this relation implies for the collinear Lagrange points, i.e., $i \in \{1,2,3\}$ that $DX_H(L_i)$ has two imaginary and two real eigenvalues.

We first need some results about 4×4-matrices belonging to the Lie algebra of the symplectic group $\mathrm{Sp}(2)$. Suppose that ω is the standard symplectic form on $\mathbb{R}^4 = \mathbb{C}^2$. Recall that a 4×4-matrix B belongs to the Lie algebra of $\mathrm{Sp}(2)$ if and only if there exists a symmetric 4×4-matrix S such that

$$B = JS$$

where J is the 4×4-matrix obtained by multiplication with i. Note that S can be recovered from B by

$$S = -JB.$$

A symmetric matrix S is called non-degenerate when it is injective. If S is non-degenerate then the *Morse index* $\mu(S) \in \{0,1,2,3,4\}$ of S is the number of negative eigenvalues of S counted with multiplicity.

Proposition 8.5.3. *Assume that* $B = JS \in \mathrm{Lie}\,\mathrm{Sp}(2)$ *such that* S *is non-degenerate and* $\mu(S) = 1$. *Then there exists a symplectic basis* $\{\eta_1, \eta_2, \xi_1, \xi_2\}$ *of* $\mathbb{R}^4$ *and* $a \in (0, \infty)$, $b \in \mathbb{R} \setminus \{0\}$ *such that*

$$B\eta_1 = a\eta_1, \quad B\eta_2 = -a\eta_2, \quad B\xi_1 = -b\xi_2, \quad B\xi_2 = b\xi_1.$$

To reformulate this statement, if $[B]_{\{\eta_1,\eta_2,\xi_1,\xi_2\}}$ denotes the matrix representation of B in the basis $\{\eta_1,\eta_2,\xi_1,\xi_2\}$ then the proposition asserts that

$$[B]_{\{\eta_1,\eta_2,\xi_1,\xi_2\}} = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & -a & 0 & 0 \\ 0 & 0 & 0 & -b \\ 0 & 0 & b & 0 \end{pmatrix}.$$

If one complexifies $\mathbb{R}^4$ to $\mathbb{C}^4$, then with respect to the basis $\{\eta_1,\eta_2,\xi_1+i\xi_2,\xi_1-i\xi_2\}$ the matrix representation of B is diagonal

$$[B]_{\{\eta_1,\eta_2,\xi_1+i\xi_2,\xi_1-i\xi_2\}} = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & -a & 0 & 0 \\ 0 & 0 & ib & 0 \\ 0 & 0 & 0 & -ib \end{pmatrix}.$$

In particular, the eigenvalues of B are $\{a,-a,ib,-ib\}$.

In order to prove Proposition 8.5.3 we need the following well-known lemma.

Lemma 8.5.4. *Assume that $B = JS \in \mathrm{Lie}\,\mathrm{Sp}(2)$ has an eigenvalue $\lambda = a+ib$ with $a,b \in \mathbb{R}\setminus\{0\}$, then the spectrum of B is*

$$\mathfrak{S}(B) = \{\lambda,-\lambda,\bar\lambda,-\bar\lambda\}.$$

Proof. Using that $J = -J^{-1} = -J^T$ and $S = S^T$ we compute

$$JBJ^{-1} = -SJ^T = -(JS)^T = -B^T$$

which shows that B is conjugated to $-B^T$. Because B and B^T have the same eigenvalues, we conclude that $-\lambda$ is an eigenvalue of B as well. Since B is real, the numbers $\bar\lambda$ and $-\bar\lambda$ are eigenvalues of B as well. By assumption neither a nor b are zero so that all the four numbers $\lambda,-\lambda,\bar\lambda,-\bar\lambda$ are different and hence the assertion about the spectrum follows. □

Proof of Proposition 8.5.3. We prove the proposition in four steps. To formulate Step 1 we abbreviate by $\mathrm{Sym}_{\mathrm{inj}}(4)$ the space of injective symmetric 4×4-matrices.

Step 1: *There exists a smooth path $S_r \in \mathrm{Sym}_{\mathrm{inj}}(4)$ for $r \in [0,1]$ with the property that*

$$S_0 = S, \quad S_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

To see that, note that because S is symmetric, there exists $R \in SO(4)$ such that $RSR^{-1} = RSR^T$ is diagonal. Moreover, because $\mu(S) = 1$ we can choose R such that the first three diagonal entries of RSR^{-1} are positive and the fourth is negative. Because $SO(4)$ is connected we can connect S and RSR^{-1} by a smooth path in $\mathrm{Sym}_{\mathrm{inj}}(4)$. Combining this path with convex interpolation between RSR^{-1} and S_1, the assertion of Step 1 follows.

Step 2: *The matrix B has two real eigenvalues and two imaginary eigenvalues.*

Let S_r for $r \in [0,1]$ be the smooth path of injective symmetric matrices obtained in Step 1. This gives rise to a smooth path $B_r = JS_r \in \mathrm{Lie}\,\mathrm{Sp}(2)$. The matrix

$$B_1 = JS_1 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

has the eigenvalues $\{1, -1, i, -i\}$. Because S_r is injective, the number 0 is not an eigenvalue of B_r for every $r \in [0,1]$. Therefore we conclude from Lemma 8.5.4 that B_r has two real and two imaginary eigenvalues for every $r \in [0,1]$.

The proof of Lemma 8.5.4 now implies that there exist $a, b \in (0, \infty)$ such that the eigenvalues of $B = B_0$ are $\{a, -a, ib, -ib\}$. In particular, there exists a basis $\{\eta_1, \eta_2, \xi_1, \xi_2\}$ of $\mathbb{R}^4$ such that

$$B\eta_1 = a\eta_1, \quad B\eta_2 = -a\eta_2, \quad B\xi_1 = -b\xi_2, \quad B\xi_2 = b\xi_1. \tag{8.26}$$

We next examine if we can choose this basis symplectic.

Step 3: *The symplectic orthogonal complement of the span of $\{\eta_1, \eta_2\}$ is spanned by $\{\xi_1, \xi_2\}$, i.e.,*

$$\langle \eta_1, \eta_2 \rangle^\omega = \langle \xi_1, \xi_2 \rangle.$$

To prove that, we first note that since $B \in \mathrm{Lie}\,\mathrm{Sp}(2)$ for every $\xi, \eta \in \mathbb{R}^4$ the formula

$$\omega(B\xi, \eta) = -\omega(\xi, B\eta)$$

holds. Hence we compute

$$b\omega(\xi_1, \eta_1) = \omega(B\xi_2, \eta_1) = -\omega(\xi_2, B\eta_1) = -a\omega(\xi_2, \eta_1) \tag{8.27}$$

and

$$b\omega(\xi_2, \eta_1) = -\omega(B\xi_1, \eta_1) = \omega(\xi_1, B\eta_1) = a\omega(\xi_1, \eta_1). \tag{8.28}$$

Combining (8.27) and (8.28) we get

$$\omega(\xi_1, \eta_1) = -\frac{a}{b}\omega(\xi_2, \eta_1) = -\frac{a^2}{b^2}\omega(\xi_1, \eta_1)$$

implying that

$$\left(1 + \frac{a^2}{b^2}\right)\omega(\xi_1, \eta_1) = 0$$

and therefore

$$\omega(\xi_1, \eta_1) = 0.$$

From (8.27) we conclude that

$$\omega(\xi_2, \eta_1) = 0$$

as well. Therefore

$$\eta_1 \in \langle \xi_1, \xi_2 \rangle^\omega$$

and the same argument with η_1 replaced by η_2 leads to

$$\eta_2 \in \langle \xi_1, \xi_2 \rangle^\omega.$$

Summarizing we showed that

$$\langle \eta_1, \eta_2 \rangle \subset \langle \xi_1, \xi_2 \rangle^\omega$$

and because ω is non-degenerate the symplectic orthogonal complement of $\langle \xi_1, \xi_2 \rangle$ is two-dimensional so that we get

$$\langle \eta_1, \eta_2 \rangle = \langle \xi_1, \xi_2 \rangle^\omega.$$

With

$$\langle \xi_1, \xi_2 \rangle = \big(\langle \xi_1, \xi_2 \rangle^\omega\big)^\omega = \langle \eta_1, \eta_2 \rangle^\omega$$

the assertion of Step 3 follows.

Step 4: *We prove the proposition.*

It follows from Step 3 that $\langle \eta_1, \eta_2 \rangle$ is a two-dimensional symplectic subspace of $\mathbb{R}^4$. Hence after rescaling η_2 we can assume that

$$\omega(\eta_1, \eta_2) = 1.$$

Note that (8.26) still remains valid after rescaling η_2. Because $\langle \xi_1, \xi_2 \rangle$ by Step 3 is a symplectic subspace as well we have

$$r := \omega(\xi_1, \xi_2) \neq 0.$$

We distinguish two cases. First assume that $r > 0$. In this case we replace ξ_1, ξ_2 by $\frac{1}{\sqrt{r}}\xi_1, \frac{1}{\sqrt{r}}\xi_2$. Then

$$\omega(\xi_1, \xi_2) = 1$$

and (8.26) still remains valid. It remains to treat the case $r < 0$. In this case we replace b by $-b$ and ξ_1, ξ_2 by $\frac{1}{\sqrt{-r}}\xi_1, -\frac{1}{\sqrt{-r}}\xi_2$. This finishes the proof of the proposition. □

Sketch of proof of Theorem 8.5.1. The collinear Lagrange points have Morse index 1. Now in view of (8.25) by Proposition 8.5.3 there exists a symplectic basis at $T_{L_i}T^*\mathbb{R}^2$ such that the linearized flow of X_H at L_i is given by

$$\phi^t = \begin{pmatrix} e^{at} & 0 & 0 & 0 \\ 0 & e^{-at} & 0 & 0 \\ 0 & 0 & \sin(bt) & \cos(bt) \\ 0 & 0 & -\cos(bt) & \sin(bt) \end{pmatrix}, \quad t \in \mathbb{R}$$

for $a \in (0, \infty)$ and $b \in \mathbb{R} \setminus \{0\}$. The result follows now by the implicit function theorem. For more details concerning this last step consult for example [110, Chapter 6] or [189, Chapter 2]. □

8.6 Sublevel sets of a Hamiltonian and 1-handles

We will revisit the connected sum arising in Hamiltonian dynamics, now that some tools are in place. In practical Hamiltonian dynamical systems, we cannot use the convenient abstract form of the connected sum we discussed in Section 6.2 of Chapter 6. Instead, we have to recognize the connected sum directly by looking at how the level sets of the Hamiltonian change. This still involves the Weinstein model. Here is the setup.

We consider a symplectic manifold (M, ω) with Hamiltonian function $H : M \to \mathbb{R}$, and make the following assumptions.

(A1) The number 0 is a critical value of H and $q_0 \in H^{-1}(0)$ is the only critical point. Furthermore this critical point is non-degenerate and $ind_{q_0} H = 1$.

(A2) There is a Liouville vector field X which is non-vanishing for all $q \in H^{-1}(c)$ for $c \leq 0$. Furthermore, for all $c \leq 0$, the vector field X is positively transverse to $H^{-1}(c) \setminus \{q_0\}$, so $X_q(H) > 0$.

(A3) Denote the quadratic form associated with the Hessian $\mathrm{Hess}_{q_0} H$ by Q. This bilinear form is non-degenerate due to the Morse assumption. Put $v = X(q_0)$. By non-degeneracy of ω and Q, we define the tangent vectors, w, w^1, w^2 by

$$Q(w, \cdot) = \omega(v, \cdot), \quad Q(w^1, \cdot) = \omega(w, \cdot), \quad Q(w^2, \cdot) = \omega(w^1, \cdot).$$

The third condition is then $\omega(v, w_2) > 0$.

Remark 8.6.1. The first condition (A1) is obviously needed for a connected sum construction, and (A2) is clearly also necessary for a contact type hypersurface. The third condition (A3) requires knowledge about the eigenvectors of the Hessian. Intuitively, this condition says the vector $X(q_0)$ should be sufficiently close to the "incoming" direction of the Weinstein model.

Denote the level set $H^{-1}(c)$ by Σ_c.

Proposition 8.6.2. *Under the above assumptions* (A1), (A2) *and* (A3), *there is* $\epsilon > 0$ *and a Liouville vector field* Z *that is positively transverse to* Σ_c *for* $0 < c < \epsilon$.

Proof. Choose a Darboux ball around $q_0 \in M$, so $\phi : B \subset (\mathbb{R}^{2n}, \omega) \to (M, \omega)$ is a local symplectomorphism mapping 0 to q_0. We get the Hamiltonian $H \circ \phi : B \to \mathbb{R}$, so we can assume that the Hamiltonian is defined on $B \subset \mathbb{R}^4$. We will do so now and just write $H : B \to \mathbb{R}$.

By the assumption that $q_0 = 0$ is a critical point of Morse-type with index 1 and the Morse lemma we see that

$$H(x) = Q(x, x) + R(x),$$

where Q is a non-degenerate quadratic form and R is smooth function of higher order, so $\lim_{x\to 0} \frac{R(x)}{|x|^2} = 0$.

According to Proposition 8.5.3 we can find a better symplectic basis with coordinates $(\eta_1, \eta_2, \xi_1, \xi_2)$ for the Darboux ball such that the Hamiltonian matrix $I \cdot Q$ attains a standard form as stated in that proposition. For the Hamiltonian Q, which is just a bilinear form, this means that

$$Q\left((\eta, \xi); (\eta, \xi)\right) = -a\eta_1\eta_2 + \frac{b}{2}(\xi_1^2 + \xi_2^2). \tag{8.29}$$

As a reminder, the symplectic form in the (η, ξ)-coordinates is given by

$$\omega = d\eta_1 \wedge d\eta_2 + d\xi_1 \wedge d\xi_2.$$

Its matrix representation will also turn out to be handy. It is given by

$$\omega(v, w) = -v^t I w := -v^t \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} w.$$

By assumption, the Liouville vector field X is non-vanishing on the entire set $H^{-1}(0)$, so by continuity it has the form $X = v \cdot \partial_x + c(x)$ near $x = 0$, where $c(x)$ is a vector field that vanishes for $x = 0$, and v is a constant non-zero vector. Define the Liouville form

$$\lambda_0 = \iota_X \omega = -v^t I \cdot dx + \gamma(x),$$

where $\gamma(x)$ is a 1-form that vanishes for $x = 0$.

We will now define another Liouville form λ_1. Consider the vector field

$$Y = \frac{1}{2}\left(\xi_1 \partial_{\xi_1} + \xi_2 \partial_{\xi_2}\right) + \left(\frac{1}{2}\eta_1 - \eta_2\right)\partial_{\eta_1} + \left(-\eta_1 + \frac{1}{2}\eta_2\right)\partial_{\eta_2}.$$

We verify that Y is a Liouville vector field. By the Cartan formula and closedness of ω we find

$$\begin{aligned} \mathcal{L}_Y \omega = d(\iota_Y \omega) &= d\left(\frac{1}{2}\eta_1 d\eta_2 - \eta_2 d\eta_2 + \eta_1 d\eta_1 - \frac{1}{2}\eta_2 d\eta_1 + \frac{1}{2}(\xi_1 d\xi_2 - \xi_2 d\xi_1)\right) \\ &= d\eta_1 \wedge d\eta_2 + d\xi_1 \wedge d\xi_2. \end{aligned}$$

Furthermore, Y is transverse to regular level sets of $x \mapsto H_Q(x) := Q(x,x)$ and of $x \mapsto H(x)$, provided the ball B is sufficiently small. Indeed, we check this directly

$$\begin{aligned}\mathcal{L}_Y(H_Q) &= \eta_1^2+\eta_2^2-\eta_1\eta_2+\frac{1}{2}\xi_1^2+\frac{1}{2}\xi_2^2\\ &= \frac{1}{2}\left((\eta_1-\eta_2)^2+\eta_1^2+\eta_2^2+\xi_1^2+\xi_2^2\right) \geq \frac{1}{2}\|x\|^2.\end{aligned}$$

Note that $\mathcal{L}_Y(H_Q) = 0$ if and only if $(\eta,\xi) = 0$. This point does not lie on a regular level of H_Q, so the first claim follows. For the second claim, note that as Y vanishes in 0, we find

$$Y(H) = Y(Q) + Y(R) \geq \frac{1}{2}\|x\|^2 + o(\|x\|^2).$$

By possibly shrinking B the claim follows.

We get the second Liouville form $\lambda_1 = \iota_Y\omega$. The difference $\lambda_1 - \lambda_0$ is closed and exact by the Poincaré lemma on a contractible open neighborhood of $x = 0$. Because we know that $\lambda_0(0) = -v^t I dx$, we find hence a function $G = -v^t I \cdot x + g(x)$ such that

$$\iota_{X+X_G}\omega = \lambda_0 - dG = \lambda_1 = \iota_Y\omega.$$

Here $g(x)$ is a function with $d_0 g = 0$, so we can and will assume that $g(x) = o(|x|)$. This motivates the definition of a new Liouville vector field

$$Z = X + X_{fG},$$

where X_{fG} is the Hamiltonian vector field associated with fG and f is a function the we shall define below. We note that for $f \equiv 1$, the Liouville vector field Z equals Y.

Construction of the separating sphere and a cutoff parameter

Since Q is a non-degenerate bilinear form, we find a unique vector w such that $Q(w,\cdot) = \omega(v,\cdot)$. In matrix language,

$$w^t Q = -v^t I, \quad w = Q^{-1} I v.$$

Define a linear function $L(x) = w^t I x$, which will serve as a cutoff parameter, and put

$$f(x) := F \circ L.$$

We have $X_L = w \cdot \partial_x$, so $X_f = F' \cdot X_L$.

We now verify that the intersection of the hyperplane $L^{-1}(0)$ and $H^{-1}(\epsilon)$ is a 2-sphere. The setup is sketched in Figure 8.4. First we consider $L^{-1}(0) \cap Q^{-1}(\epsilon)$. Note that $\ker L = \operatorname{span}(w, Jw, Kw)$, where

$$J = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

together with I, these matrices represent the standard imaginary quaternions.

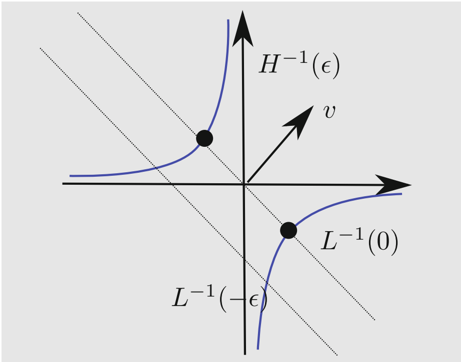


Figure 8.4: The hyperplane $L = 0$ and hypersurface $H^{-1}(\epsilon)$ intersecting in a 2-sphere.

We now give the matrix representation of $Q|_{\ker L}$ with respect to this basis $w_1 = w, w_2 = Jw, w_3 = Kw$. This is done with a tedious computation consisting of the six distinct products $w_i^t Q w_j$ with $i \leq j$. To get somewhat nicer looking formulas, we will write

$$v = (v_1, v_2, v_3, v_4) = (v_{\eta_1}, v_{\eta_2}, v_{\xi_1}, v_{\xi_2}) = (v_\eta, v_\xi).$$

The matrix representation $[Q|_{\ker L}]$ can now directly computed, but the expression is rather long, so we will tensor notation, meaning that $[Q|_{\ker L}]$ is written in the form $[Q|_{\ker L}] = [Q|_{\ker L}]_{ij} e_i \otimes e_j$. We find

$$\begin{aligned}
[Q|_{\ker L}] = {} & \frac{2v_{\eta_1}v_{\eta_2}b + \|v_\xi\|^2}{b} e_1 \otimes e_1 + \frac{{v_{\eta_1}}^2 b^3 + {v_{\eta_2}}^2 b^3 - 2v_{\xi_2}v_{\xi_1}}{b^2} e_2 \otimes e_2 \\
& + \frac{{v_{\eta_1}}^2 b^3 + {v_{\eta_2}}^2 b^3 + 2v_{\xi_2}v_{\xi_1}}{b^2} e_3 \otimes e_3 + 2\frac{{v_{\xi_1}}^2 - {v_{\xi_2}}^2}{b^2} e_2 \otimes e_3 \\
& + 2\frac{(v_{\xi_2}v_{\eta_1} - v_{\xi_1}v_{\eta_2})b + v_{\xi_1}v_{\eta_1} - v_{\xi_2}v_{\eta_2}}{b} e_1 \otimes e_2 \\
& + 2\frac{v_{\xi_2}v_{\eta_1} + v_{\xi_1}v_{\eta_2} - (v_{\xi_1}v_{\eta_1} + v_{\xi_2}v_{\eta_2})b}{b} e_1 \otimes e_3.
\end{aligned}$$

Its determinant is given by

$$\begin{aligned}
\det[Q|_{\ker L}] &= \frac{1}{b^5}\left(2v_1v_2\|v_\eta\|^4 b^7 + 4v_1v_2\|v_\xi\|^2\|v_\eta\|^2 b^5 - \|v_\xi\|^2|v_\eta\|^4 b^4\right. \\
&\qquad \left. + 2v_1v_2\|v_\xi\|^4 b^3 - 2\|v_\xi\|^4\|v_\eta\|^2 b^2 - \|v_\xi\|^6\right) \\
&= \frac{1}{b^5}\left(\|v_\eta\|^2 b^2 + \|v_\xi\|^2\right)^2 \left(2v_1v_2b^3 - \|v_\xi\|^2\right)
\end{aligned}$$

Due to the final factorization, this is obviously positive precisely when $2b^3 v_{\eta_1} v_{\eta_2} - \|v_\xi\|^2 > 0$. The following lemma tells us that this is equivalent to condition (A3).

Lemma 8.6.3. *In the symplectic coordinates such that Q attains the standard form* (8.29)*, condition* (A3) *is equivalent to* $2b^3 v_{\eta_1} v_{\eta_2} - \|v_\xi\|^2 > 0$.

Proof. In coordinates, the equations for w, w^1 and w^2 become

$$w = Q^{-1}Iv, \quad w^1 = Q^{-1}IQ^{-1}Iv, \quad w^2 = Q^{-1}IQ^{-1}IQ^{-1}Iv,$$

where I is the standard matrix representation for $-\omega$. We write out condition (A3) and find

$$\begin{aligned}
\omega(v, w_2) &= -v^t IQ^{-1}IQ^{-1}IQ^{-1}Iv \\
&= -v^t \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & b^{-1} \\ 0 & 0 & -b^{-1} & 0 \end{pmatrix}^3 v \\
&= 2v_{\eta_1} v_{\eta_2} - \frac{\|v_\xi\|^2}{b^3}.
\end{aligned}$$

This proves the claim. □

We continue the construction of the separating sphere. We claim that if condition (A3) holds, then the matrix $[Q|_{\ker L}]$ is positive definite. To see this, consider the homotopy $v(s) = (v_\eta, (1-s)v_\xi)$. If $v(0)$ satisfies $2v_1 v_2 b^3 - \|v_\xi\|^2 > 0$, then so does $v(s)$ for all $s \in [0,1]$, so $[Q|_{\ker L}]$ is non-degenerate for all $s \in [0,1]$. For $s = 1$, all off-diagonal terms in $[Q|_{\ker L}]$ drop out and the entries on the diagonal are all positive, so $v(1)$ is positive definite, and hence so are the $v(s)$. We conclude that the quadric $L^{-1}(0) \cap Q^{-1}(\epsilon)$ in the three-dimensional vector space $L^{-1}(0)$ is a 2-sphere. It now also follows that $L^{-1}(0) \cap H^{-1}(\epsilon)$ is also diffeomorphic to a 2-sphere provided we choose ϵ sufficiently small.

Claim: The vector field Z is everywhere transverse to $H^{-1}(\epsilon)$.
Using that $X_{fG} = GX_f + fX_G$, we compute

$$\begin{aligned}
Z(H) &= X(H) + fX_G(H) + GX_f(H) \\
&= (1-f)X(H) + f(X + X_G)(H) + GX_f(H) \\
&= (1-f)X(H) + fY(H) + F'GX_L(H) \\
&= (1-f)X(H) + fY(H) + F'(-v^t Ix + g(x))(2Q(w,x) + o(|x|)) \\
&= (1-f)X(H) + fY(H) + F'(2(v^t Ix)^2 + o(|x|^2)).
\end{aligned}$$

By choosing ϵ sufficiently small, we can ensure that x is small in norm making the $o(|x|^2)$-term smaller than $Y(H) \geq \frac{1}{4}\|x\|^2$.

We now complete the claim that Z is transverse to Σ_c. There are two cases to consider.

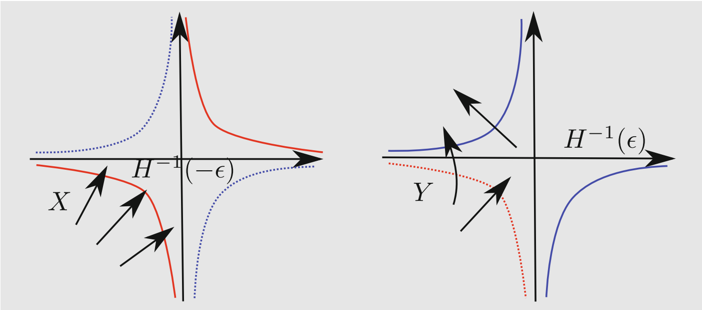


Figure 8.5: Liouville vector field near a critical point.

- Away from the Darboux ball $\phi(B)$, the vector field Z coincides with the original vector field X. Furthermore, for sufficiently small $c > 0$ and points x not in $\phi(B)$, this Liouville vector field X was assumed to be transverse to Σ_c.
- In the Darboux ball $\phi(B)$, each of the terms $(1-f)X(H)$, $fY(H)$ and $-2F'(v^tIx)^2)$ are non-negative. Furthermore, if $f = 0$ (this is away from the hyperplane $L^{-1}(0)$) the first term is positive by assumption. If $f = 1$, the second term is positive by our earlier computation, and for values of f between 0 and 1 the first and second term are positive and their sum dominates the last term. □

As an example, we apply this proposition to Hill's lunar problem. In that case, the matrix Q in (q, p)-coordinates is given by the Hessian of Hill's lunar Hamiltonian, so by Equation (5.29) we find at the critical points

$$Q = \begin{pmatrix} -9 & 0 & 0 & 1 \\ 0 & 3 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Writing out $\omega(v, w_2)$ we find

$$\omega(v, w_2) = \frac{3}{20},$$

so the proposition applies in the bounded component of Hill's lunar problem.

Remark 8.6.4. This does not yet show that Hill's lunar problem is of contact type above the first critical value, since we haven't constructed a transverse Liouville vector field in the unbounded component.

On the hand, this proposition can be applied to the two bounded components in the restricted three-body problem giving an alternative argument for connected sum in [9].

Chapter 9

Global Surfaces of Section

9.1 Disk-like global surfaces of section

Assume that $X \in \Gamma(TS^3)$ is a non-vanishing vector field on the three-dimensional sphere S^3. We denote by ϕ_X^t its flow on S^3.

Definition 9.1.1. A (*disk-like*) *global surface of section* is an embedded disk $D \subset S^3$ satisfying

(i) X is tangent to ∂D, the boundary of D,

(ii) X is transverse to the interior $\mathring{D}$ of D,

(iii) For every $x \in S^3 \setminus \partial D$ there exists $t^+ > 0$ and $t^- < 0$ such that $\phi_X^{t^+}(x) \in \mathring{D}$ and $\phi_X^{t^-}(x) \in \mathring{D}$.

Remark 9.1.2. Requirement (i) implies that the boundary ∂D of the disk is a periodic orbit of X. We refer to ∂D as the *binding orbit* of the global surface of section. In the construction of global surfaces of section via holomorphic curves, the binding orbit corresponds to the binding of an open book, see Section 17.1.

In Figure 9.1 we illustrate the concept of a global surface of section in a simplified setting, where the ambient space is S^2 and the "surface" is the segment connecting the two fixed points of a flow on S^2. Except for the fixed points each orbit hits the segment both in forward and backward time.

Remark 9.1.3. Instead of a disk one could consider more generally a Riemann surface with boundary. In particular, an important example is an annulus which has two binding orbits, see Figure 9.2. However, in the following we concentrate ourselves on disks and mean by a global surface of section always a disk-like global surface of section unless specified otherwise.

Let us now assume that $D \subset S^3$ is a global surface of section. We define the *Poincaré return map*

$$\psi\colon \mathring{D} \to \mathring{D}$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_9

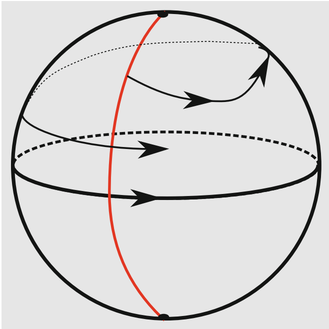


Figure 9.1: A global segment of section.

as follows. Given $x \in \mathring{D}$, define

$$\tau(x) := \min\left\{t > 0 : \phi_X^t(x) \in \mathring{D}\right\},$$

i.e., the next return time of x to D. It follows from the conditions of a global surface of section that $\tau(x)$ exists and is finite. Moreover, the function

$$\tau\colon \mathring{D} \to (0,\infty), \quad x \mapsto \tau(x)$$

is smooth. Now define

$$\psi(x) := \phi_X^{\tau(x)}(x).$$

If $x \in \mathring{D}$ and $\xi \in T_xD$, the differential of the Poincaré return map is given by

$$d\psi(x)\xi = d\phi_X^{\tau(x)}(x)\xi + (d\tau(x)\xi)X. \tag{9.1}$$

The two-dimensional disk D together with the Poincaré return map ψ basically contains all the relevant information on the flow of X on the three-dimensional manifold S^3. Instead of the continuous flow on the three-dimensional manifold we can study alternatively the discrete flow of the Poincaré return map, so that we can think of a global surface of section as a stroboscope. For example, periodic orbits of X different from the binding orbit ∂D correspond to periodic points of the Poincaré return map. One can say that a global surface of section reduces the complexity of the problem by one dimension.

Recall that a Hamiltonian structure on S^3 is a closed two-form $\omega \in \Omega^2(S^3)$ with the property that $\ker\omega$ is a one-dimensional distribution. A non-vanishing section $X \in \Gamma(\ker\omega)$ is referred to as a Hamiltonian vector field. In view of (9.1) the following Lemma follows.

Lemma 9.1.4. *Assume that $\omega \in \Omega^2(S^3)$ is a Hamiltonian structure, $X \in \Gamma(\ker\omega)$, and D is a global surface of section for X with Poincaré return map ψ. Then*

$$\psi^*\omega|_{\mathring{D}} = \omega|_{\mathring{D}},$$

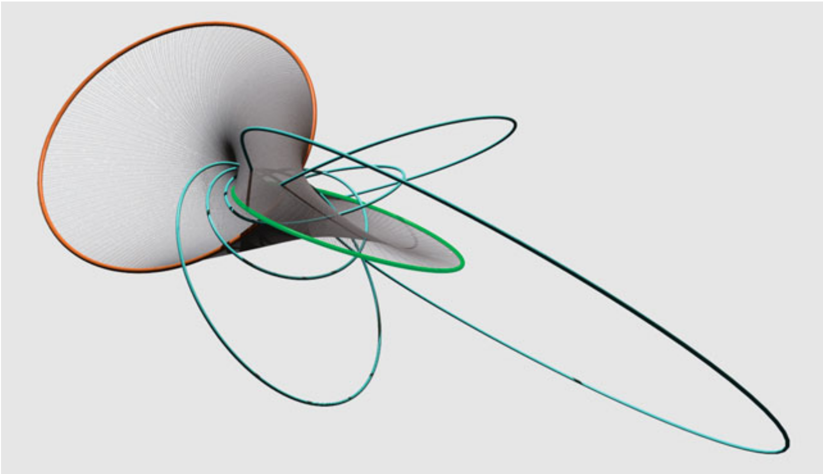


Figure 9.2: The existence of an *annulus-like* surface of section has been known in the component of the heavy primary for small μ since Poincaré and Birkhoff. This figure displays such a surface after stereographic projection of the Levi-Civita regularization. Circle-shaped orbit on the left is the retrograde orbit, and the small circular orbit on the right is the direct orbit. In addition, we draw one extra orbit and its intersections with the annulus.

i.e., ψ is area preserving with respect to the restriction of ω to the interior of the global surface of section.

Remark 9.1.5. The concept of a global surface of section basically already appears in Section 305 of Chapter XXVII of Poincaré's "Nouvelles Méthodes de la mécanique céleste" [205]. In that section, Poincaré describes an orbit intersecting a half-plane infinitely many times. See the excellent article of Chenciner [54, Section 9] for enlightening discussions of what Poincaré does in this section. Shortly before his death, Poincaré also describes an annulus-type global surface of section and some of its dynamical properties in [207]. In this paper, Poincaré formulated a conjecture referred to as his "Last Geometric Theorem" concerning annulus maps. This conjecture asserts the existence of infinitely many periodic orbits provided that there is an annulus-type global surface of section whose return map satisfies a certain twist condition.

Shortly after the death of Poincaré this conjecture was proved by Birkhoff [38]. Poincaré had the vision that the annulus type global surface of section has as binding orbits the retrograde and the direct one. For small energy such an annulus type global surface of section actually exists as was shown by Conley [57] and Kummer [155]. However, Birkhoff pointed out in [39] that the direct periodic orbit becomes degenerate below the first critical value. He therefore introduced the concept of a disk-like global surface of section and conjectured that below the first critical value in the bounded component there always exists a disk-like global surface of section for the restricted three-body problem with as binding orbit the retrograde orbit.

In Figure 9.2 we display an annulus-like surface of section in the restricted three-body problem, which we obtained through numerical means.

Figure 9.3: Iterates of some orbits in the annular surface of section for $\mu = 0.539$ and $c = 2.0$. See the final chapter for an explanation.

9.2 Obstructions

Given a non-vanishing vector field X on S^3 it is far from obvious that the dynamical system (S^3, X) admits a global surface of section. Moreover, given a periodic orbit γ of the vector field X we would like to know if it bounds a global surface of section. A periodic orbit can be interpreted as a knot in S^3, i.e., an isotopy class of embeddings of $S^1 \to S^3$. The first obstruction is obvious.

Obstruction 1. If a periodic orbit is a binding orbit of a global surface of section, then it is unknotted.

Indeed, since the periodic orbit is the boundary of an embedded disk it has Seifert genus zero. This is a characterizing property of the unknot. The second obstruction describes the relation of the binding orbit with all other periodic orbits.

Obstruction 2. If a periodic orbit is the binding orbit of a global surface of section, it is linked to every other periodic orbit.

Indeed, a periodic orbit different from the binding orbit is a periodic point of the Poincaré return map, i.e., coincides with a fixed point of an iteration of the Poincaré return map. Because X is transverse to the interior of the global surface of section, each intersection point of the periodic orbit with the global surface of section counts with the same sign and therefore the linking number does not vanish.

There is a third obstruction which requires us to explain the concept of a *self-linking number*. For the notion of self-linking number assume that the vector field X is the Reeb vector field of a contact form λ on S^3. Abbreviate by $\xi = \ker \lambda$ the hyperplane distribution in T^*S^3. In contact geometry there are two important classes of knots. A *Legendrian knot* is an embedding $\gamma\colon S^1 \to S^3$ with the property that $\partial_t \gamma(t) \in \xi_{\gamma(t)}$ for every $t \in S^1$. The other extreme is a *transverse knot*, which is an embedding $\gamma\colon S^1 \to S^3$ with the property that $\partial_t \gamma(t) \notin \xi_{\gamma(t)}$ for every $t \in S^1$. We refer to the book by Geiges [97] for a detailed discussion of Legendrian and transverse knots in contact topology. A periodic Reeb orbit is an example of a transverse knot and therefore we restrict our discussion in the following to transverse knots.

We explain now the notion of self-linking number for a transverse unknot $\gamma\colon S^1 \to S^3$. Abbreviating by $D = \{z \in \mathbb{C} : |z| \leq 1\}$ the unit disk in $\mathbb{C}$ we first choose an embedding

$$\overline{\gamma}\colon D \to S^3$$

with the property that

$$\overline{\gamma}(e^{2\pi i t}) = \gamma(t).$$

That such an embedding exists follows from the assumption that γ is the unknot. Consider the vector bundle of rank two $\overline{\gamma}^*\xi \to D$. Because D is contractible, we can choose a non-vanishing section $X\colon D \to \overline{\gamma}^*\xi$. Fix a Riemannian metric g on S^3 and define

$$\gamma_X\colon S^1 \to S^3, \quad t \mapsto \exp_{\gamma(t)} X(t).$$

Because γ is a transverse knot we can choose X so small such that

$$\gamma_{rX} \cap \gamma = \emptyset, \quad r \in (0,1].$$

We define the self-linking number of γ to be

$$s\ell(\gamma) := \ell k(\gamma, \gamma_X) \in \mathbb{Z}$$

where ℓk is the linking number. By homotopy invariance of the linking number the self-linking number of γ does not depend on the choice of the section X and the Riemannian metric g. It is also independent of the choice of the embedded filling disk $\overline{\gamma}$. To see that note that if $\overline{\gamma}\colon D \to S^3$ and $\overline{\gamma}'\colon D \to S^3$ are two embedded filling disks of γ in view of the fact that $\pi_2(S^3) = \{0\}$ the two filling disks are homotopic. Even if the homotopy is not through embedded disks we can use it to construct for a given non-vanishing section $X\colon D \to \overline{\gamma}^*\xi$ a non-vanishing section

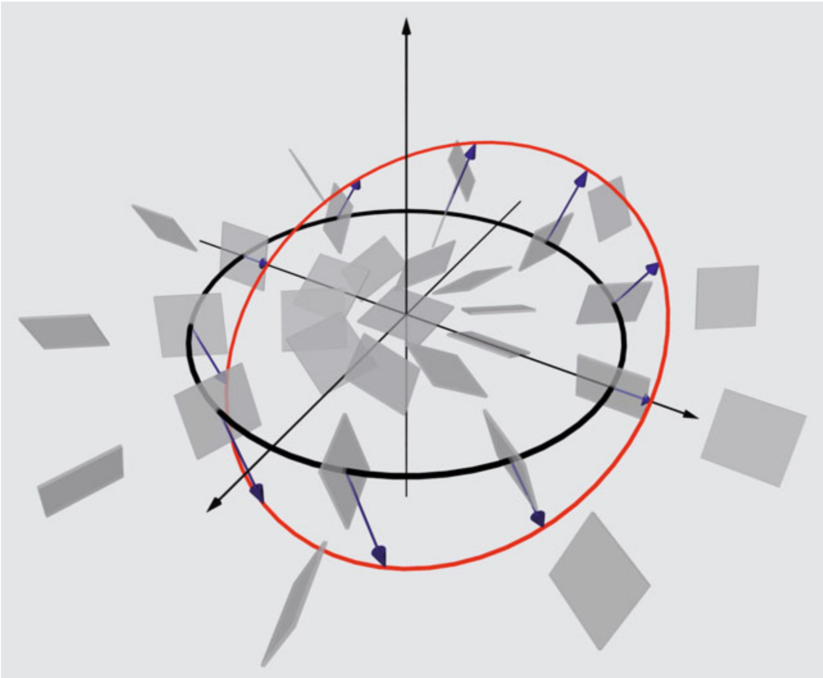


Figure 9.4: A transverse unknot γ and its pushoff γ_X.

$X' \colon D \to (\overline{\gamma}')^*\xi$ with the property that the restrictions of the two sections to the boundary $\partial D = S^1$ satisfy

$$X|_{\partial D} = X'|_{\partial D} \colon S^1 \to \gamma^*\xi.$$

This shows that the self-linking number is independent of the choice of the embedded filling disk as well.

Remark 9.2.1. The self-linking number can also be defined for transverse knots $\gamma \colon S^1 \to S^3$ which are not necessarily unknots. One uses here the fact that for every knot there exists an oriented surface with boundary Σ and an embedding $\overline{\gamma} \colon \Sigma \to S^3$ with the property that $\overline{\gamma}|_{\partial\Sigma} = \gamma$. Such a surface is referred to as a *Seifert surface*, see for example [162], and with the help of a Seifert surface one defines the self-linking number of a transverse knot similarly as for unknots. We refer to the book of Geiges [97] for details. For the unknot the Seifert surface can be chosen as a disk and if one defines the genus of a knot to be

$$g(\gamma) := \min\big\{g(\Sigma) : \Sigma \,\text{Seifert surface of}\ \gamma\big\} \in \mathbb{N} \cup \{0\},$$

then the unknot can be characterized as the knot whose genus vanishes.

We are now ready to formulate the third obstruction for a periodic orbit to be the binding orbit of a global surface of section.

Obstruction 3 If the vector field on S^3 coincides with the Reeb vector field of a contact form on S^3, and a periodic Reeb orbit γ is the binding orbit of a global surface of section, then its self-linking number satisfies $s\ell(\gamma) = -1$.

We next explain the reason for Obstruction 3. Let $D \subset S^3$ be a global surface of section. We then have two rank-2 vector bundles over the disk $\xi|_D \to D$ and $TD \to D$. Abbreviate by

$$\pi\colon TS^3 \to \xi$$

the projection along the Reeb vector field R. Since R is transverse to the interior of D we obtain a bundle isomorphism

$$\pi|_{\mathring{D}}\colon T\mathring{D} \to \xi|_{\mathring{D}}.$$

Choose

$$X\colon D \to \xi$$

a non-vanishing section. If (r, θ) are polar coordinates on D, we define another section from D to ξ by

$$Y := \pi(r\partial_r)\colon D \to \xi.$$

Because ∂D is a periodic Reeb orbit, it holds that $\partial_\theta|_{\partial D}$ is parallel to the Reeb vector field and therefore $Y|_{\partial D}\colon \partial D \to \xi$ is non-vanishing. In particular, we have two non-vanishing section $X|_{\partial D}, Y|_{\partial D}\colon \partial D \to \xi$. The self-linking number of the Reeb orbit γ is then given by

$$s\ell(\gamma) = s\ell(\partial D) = \mathrm{wind}_{\partial D}(Y, X) = -\mathrm{wind}_{\partial D}(X, Y),$$

where $\mathrm{wind}_{\partial D}(Y, X)$ is the winding number of X around Y computed with respect to the orientation of ξ induced by $d\lambda$. For $r \in (0, 1]$ abbreviate by $D_r = \{z \in D : |z| \leq r\}$ the ball of radius r. The section X is non-vanishing on D while the section Y only vanishes at 0. Therefore the winding number

$$\mathrm{wind}_{\partial D_r}(Y, X) \in \mathbb{Z}$$

is defined for every $r \in (0, 1]$. By homotopy invariance of the winding number we conclude that

$$\mathrm{wind}_{\partial D}(Y, X) = \mathrm{wind}_{\partial D_r}(Y, X), \quad \forall\, r \in (0, 1].$$

We now look at the situation for $r = \delta$ close to 0. Since $X(0)$ is non-vanishing, we see that $Y = \pi(r\partial_r)$ winds once around $X(0)$ and therefore

$$\mathrm{wind}_{\partial D_\delta}(X, Y) = 1.$$

Combining these facts we conclude that

$$s\ell(\gamma) = -1.$$

We point out that to derive Obstruction 3 we only used the local assumption (i) and (ii) of Definition 9.1.1 and not the global assumption (iii).

9.3 Perturbative methods

In this section we describe McGehee's construction of a global disk-like surface of section and its return map following McGehee's thesis, [174]. We start with the Hamiltonian for the rotating Kepler problem using complex coordinates. The Hamiltonian is then defined by

$$H = \frac{1}{2}p\bar{p} - \frac{i}{2}(q\bar{p} - \bar{q}p) - \frac{1}{\sqrt{q\bar{q}}}.$$

This is defined on $T^*\mathbb{C}^*$ with the symplectic form

$$\omega_0 = \frac{1}{2}\left(dp \wedge d\bar{q} + d\bar{p} \wedge dq\right).$$

We perform the Levi-Civita regularization on the energy level $H = -c$, where $c > 3/2$. In other words, substitute

$$q = 2z^2, \quad p = \frac{w}{\bar{z}},$$

and define

$$K := (H + c)|z|^2 = \frac{1}{2}\left(w\bar{w} + z\bar{z} \cdot 2 \cdot (c - i(z\bar{w} - \bar{z}w)) - 1\right).$$

We recall that the angular momentum is preserved in the rotating Kepler problem, so it makes sense to make the identification $E - L = -c$, where the angular momentum L is defined by

$$L = \frac{i}{2}(q\bar{p} - \bar{q}p) = i(z\bar{w} - \bar{z}w).$$

In particular we see that

$$2 \cdot (c - i(z\bar{w} - \bar{z}w)) = -2E. \tag{9.2}$$

The Levi-Civita transformation pulls back the standard form ω_0 to

$$\frac{1}{2}\left(d\left(\frac{w}{\bar{z}}\right) \wedge d(2\bar{z}^2) + d\left(\frac{\bar{w}}{z}\right) \wedge d(2z^2)\right) = 2\left(dw \wedge d\bar{z} + d\bar{w} \wedge dz\right) = 4\omega,$$

where ω is the standard symplectic form. To find the equations of motion, we continue to use the standard form, since this will give the same Hamiltonian vector field up to a factor 4. We find the following equations of motion

$$\begin{aligned} \dot{z} &= 2\frac{\partial K}{\partial \bar{w}} = w - 2iz|z|^2 \\ \dot{w} &= -2\frac{\partial K}{\partial \bar{z}} = 2Ez - 2iw|z|^2. \end{aligned} \tag{9.3}$$

In the last step we have used the above identification (9.2). At this point we stop using the Hamiltonian description, and we will make the non-canonical coordinate transformation, which was also used by Conley in [57],

$$\begin{aligned}\zeta_1 &= w + i\sqrt{-2E}z \\ \zeta_2 &= \bar{w} + i\sqrt{-2E}\bar{z}.\end{aligned}$$

Lemma 9.3.1. *The points* (ζ_1, ζ_2) *lie on a sphere with radius* $\sqrt{2}$*.*

Proof. We compute

$$\begin{aligned}|\zeta_1|^2 + |\zeta_2|^2 &= |w|^2 - 2E|z|^2 + i\sqrt{-2E}(z\bar{w} - \bar{z}w) \\ &\quad + |w|^2 - 2E|z|^2 - i\sqrt{-2E}(z\bar{w} - \bar{z}w) \\ &= 2(|w|^2 - 2E|z|^2) = 2.\end{aligned}$$

The last step follows as we are looking at the energy surface $K = 0$, which is a preserved quantity. □

We write out the equations of motions for ζ_1 and ζ_2 by simply substituting (9.3). We find

$$\begin{aligned}\dot{\zeta}_1 &= \dot{w} + i\sqrt{-2E}\dot{z} = 2Ez - 2iw|Z|^2 + i\sqrt{-2E}w + \sqrt{-2E}2z|z|^2 \\ &= i(\sqrt{-2E} - 2|z|^2)\zeta_1 \\ \dot{\zeta}_2 &= \dot{\bar{w}} + i\sqrt{-2E}\dot{\bar{z}} = i(\sqrt{-2E} + 2|z|^2)\zeta_2.\end{aligned} \tag{9.4}$$

We make one more coordinate change to make the flow more uniform. Define

$$\psi := \frac{z\bar{w} + \bar{z}w}{-4E} = \frac{2\,\mathrm{re}(z\bar{w})}{-4E},$$

and put

$$\xi_1 := e^{-i\psi}\zeta_1, \quad \xi_2 := e^{i\psi}\zeta_2.$$

To find the equations of motion we need to compute

$$\begin{aligned}\dot{\psi} &= \frac{2\,\mathrm{re}(\dot{z}\bar{w} + z\dot{\bar{w}})}{-4E} = \frac{2\,\mathrm{re}(w\bar{w} - z\bar{w}2i|z|^2 + 2E\bar{z}z + 2i\bar{w}z|z|^2)}{-4E} \\ &= \frac{|w|^2 + 2E|z|^2}{-2E} = \frac{1 + 4E|z|^2}{-2E},\end{aligned}$$

where we have used in the last step that $K = 0$. The new equations of motion are hence

$$\begin{aligned}\dot{\xi}_1 &= i\left(\sqrt{-2E} - 2|z|^2 - \dot{\psi}\right)\xi_1 = i\left(\sqrt{-2E} - \frac{1}{(\sqrt{-2E})^2}\right)\xi_1 \\ \dot{\xi}_2 &= i\left(\sqrt{-2E} + 2|z|^2 + \dot{\psi}\right)\xi_2 = i\left(\sqrt{-2E} + \frac{1}{(\sqrt{-2E})^2}\right)\xi_2.\end{aligned} \tag{9.5}$$

Remark 9.3.2. As before we have $|\xi_1|^2+|\xi_2|^2=2$, so the flow has been regularized to a standard copy of S^3 with radius $\sqrt{2}$. Furthermore, the arguments (i.e., phases) of ξ_1 and ξ_2 rotate with constant speed $\sqrt{-2E}-\frac{1}{(\sqrt{-2E})^2}$ and $\sqrt{-2E}+\frac{1}{(\sqrt{-2E})^2}$, respectively.

There are two obvious periodic orbits, namely one with $\xi_1=0$ and one with $\xi_2=0$.

Lemma 9.3.3. *The orbit with $\xi_2=0$ is the retrograde, circular orbit, which we will denote by γ_{retro}, whereas the orbit with $\xi_1=0$ is the direct circular orbit.*

Proof. An orbit in the Kepler problem is circular precisely when its eccentricity vanishes, so the Runge–Lenz vector satisfies $A=0$. Hence according to Equation (3.7) an orbit is circular if and only if $1+2EL^2=0$. The orbits with $\xi_1=0$ or $\xi_2=0$ satisfy this equation. For instance, with the above formula for L we compute

$$0=|\xi_1|^2=|w|^2-2E|z|^2+\sqrt{-2E}i(z\bar{w}-\bar{z}w)=1+\sqrt{-2E}L,$$

whence $-2EL^2=1$. Similarly, we find for $\xi_2=0$ that $1-\sqrt{-2E}L$, which also implies $-2EL^2=1$. Furthermore, both orbits with $\xi_1=0$ or $\xi_2=0$ are in the bounded component of the Hill's region. To see that the orbit with $\xi_1=0$ is direct, we observe that L must be negative. For the orbit with $\xi_2=0$, we see that L is positive, so that orbit is retrograde. □

9.3.1 Global surface of section

We now get explicit global surfaces of section

$$\begin{aligned} D_{\mathrm{retro},\theta}: D^2_{\sqrt{2}} &\longrightarrow S^3_{\sqrt{2}} \\ (r,\phi) &\longmapsto \left(re^{i\phi};\sqrt{2-r^2}e^{i\theta}\right). \end{aligned} \tag{9.6}$$

In other words, the interior of $D_{\mathrm{retro},\theta}$ is the fiber of the map

$$S^3-\gamma_{\mathrm{retro}}\longrightarrow S^1,\quad (z_1,z_2)\longmapsto\frac{z_2}{|z_2|}.$$

The following proposition is now clear

Proposition 9.3.4. *Each disk $D_{\mathrm{retro},\theta}$ is a global disk-like surface of section whose boundary is the retrograde, circular orbit.*

Proof. We check all properties for the disk $D_{\mathrm{retro},\theta}$

1. The boundary of $D_{\mathrm{retro},\theta}$ satisfies $\xi_2=0$, so it is the retrograde circular orbit by Lemma 9.3.3.

2. That the flow is transverse to the interior of the disk follows from Remark 9.3.2: the argument of ξ_2 is strictly increasing with derivative $v(E) = \sqrt{-2E} + \frac{1}{(\sqrt{-2E})^2}$,
3. Similarly, we also see that all orbits must come back, since we find a return time equal to $2\pi/v(E)$. □

Remark 9.3.5. From the global surface of section and an explicit formula for the return map, obtained from Equation (9.5), McGehee obtains a lot of interesting, dynamical information. Obviously, many invariant curves are visible, and the twist condition of KAM theory can be verified. KAM theory, short for Kolmogorov–Arnold–Moser, is a theory that asserts the persistence of quasiperiodic orbits under small perturbations. For a brief overview, see the book of Hall and Meyer, [110], and the encyclopedic work by Arnold, Kozlov and Neishtadt, [20].

An important upshot of McGehee's work is that he proves that the Earth (in the Earth-Moon system for a light Moon) is protected by invariant tori; one cannot travel from a neighborhood of the retrograde orbit to the Moon without using fuel.

9.4 Existence results from holomorphic curve theory

Before we describe the general results, we start with a simple explicit example where many of the general features are already present.

9.4.1 A simple example

Consider S^3 with the standard contact form $\alpha = \frac{i}{2}\sum_j (z_j d\bar{z}_j - \bar{z}_j dz_j)|_{S^3}$. This form can be obtained in several ways. Define the function $f : \mathbb{C}^2 \to \mathbb{R}$ by sending (z_1, z_2) to $\frac{1}{2}(|z_1|^2 + |z_2|^2)$.

- $\lambda = -df \circ i$.
- $\lambda = \iota_X \omega$ with $X = \frac{1}{2}(z \cdot \partial_z + \bar{z} \cdot \partial_{\bar{z}})$.

These are equivalent. The contact form α is obtained by restricting to S^3.

Lemma 9.4.1. *The map*

$$\begin{aligned} \psi : (\mathbb{R} \times S^3, d(e^t \alpha)) &\longrightarrow (\mathbb{C}^2 \setminus \{0\}, \omega_0) \\ (t, p) &\longmapsto e^{t/2} j_{S^3}(p) \end{aligned}$$

is a symplectomorphism. Furthermore, the standard complex structure on $\mathbb{C}^2 \setminus \{0\}$ defines an SFT-like almost complex structure under this identification.

By i being an SFT-like almost complex structure, we mean that i is invariant under the Liouville flow of X, sends the Liouville vector field to the Reeb vector field, and sends the contact structure $\xi = \ker \alpha$ to ξ. We will see this notion in more detail in Section 13.1.

Proof. The first statement can be proved by a direct check or by observing that the map sends Liouville flow lines of the Liouville vector field ∂_t on $\mathbb{R} \times S^3$ to Liouville flow lines of the Liouville vector field X on $\mathbb{C}^2 \setminus \{0\}$.

To see that the complex structure is of SFT-type, we work on $\mathbb{C}^2 \setminus \{0\}$. We have

- $iX = R$, and $i\xi = \xi$.
- to see that i is translation invariant, we use standard coordinates and note that the Liouville flow just rescales a point $p \in \mathbb{C}^2$ by a factor $e^{t/2}$. Hence we have

$$d\phi_X^t{}^{-1} \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} d\phi_X^t = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$ □

We now give a very quick preview of holomorphic curves in symplectizations. With the standard complex structure on $\mathbb{C}^2 \setminus \{0\}$ we can use basic complex analysis to write down holomorphic curves. These can be translated back into the symplectization setting using the inverse of ψ, which is given by

$$\begin{aligned} \psi^{-1} := (\pi_{\mathbb{R}}, \pi_{S^3}) : \mathbb{C}^2 \setminus \{0\} &\longrightarrow \mathbb{R} \times S^3, \\ (z_1, z_2) &\longmapsto \left(\log(|z_1|^2 + |z_2|^2); \frac{(z_1, z_2)}{\sqrt{|z_1|^2 + |z_2|^2}} \right). \end{aligned}$$

We will restrict ourselves to genus 0-curves, or in other words, curves whose domain has the form $\mathbb{C} \setminus \{q_1, \dots, q_k\}$. We can now take any pair of meromorphic functions f and g, and define a map of the form

$$u : \mathbb{C} \setminus \{q_1, \dots, q_k\} \longrightarrow \mathbb{C}^2 \setminus \{0\}, \quad z \longmapsto (f(z), g(z)).$$

However, if these functions have an essential singularity at infinity, then the behavior near infinity will be wild. For example, the theorem of Casorati–Weierstrass tells us that $\pi_{\mathbb{R}} \circ u$ will oscillate between $-\infty$ and ∞ in every neighborhood of infinity.

If we restrict ourselves to rational functions f and g, we get much better behavior. In that case, near each puncture z_0 the map $\pi_{\mathbb{R}} \circ u(z_0 + \delta e^{it})$ converges to either $-\infty$ or ∞ as $\delta \to 0$ (or $\delta \to \infty$ for the puncture at infinity). Furthermore, we get a relation with the dynamics as the map $\pi_{S^3} \circ u(z_0 + \delta e^{it})$ converges to a covering of a Hopf fiber. Of course, Hopf fibers are precisely the periodic Reeb orbits for this special contact form. In general, as we will see in later sections, the relation between holomorphic curves and the Reeb dynamics comes from the special choice of almost complex structure. Furthermore, wild behavior can be excluded by considering curves with finite *Hofer energy*. We will define this notion in Section 13.2.

Foliation by holomorphic planes

Fix $a \in \mathbb{C}^*$, and define $u_a : \mathbb{C} \to \mathbb{C}^2 \setminus \{0\}$ by

$$u_a(z) = (z, a).$$

Together with the so-called orbit cylinder $u_0 : \mathbb{C}^* \to \mathbb{C}^2 \setminus \{0\}$, these curves foliate $\mathbb{C}^2 \setminus \{0\}$. For the asymptotic convergence we use cylindrical coordinates for both the domain $\mathbb{C}$ and the target $\mathbb{C}^2 \setminus \{0\} \cong \mathbb{R} \times S^3$,

$$\begin{aligned} c : \mathbb{R} \times S^1 &\longrightarrow \mathbb{C} \\ (r, \phi) &\longmapsto e^{r+i\phi}. \end{aligned}$$

Then we find

$$\begin{aligned} U_a(r, \phi) &= \psi^{-1} \circ u_a \circ c(r, \phi) \\ &= \left(\log\left(e^{2r} + |a|^2\right); \frac{e^{i\phi}}{\sqrt{1 + e^{-2r}|a|^2}}, \frac{a}{\sqrt{1 + e^{-2r}|a|^2}} \right). \end{aligned} \tag{9.7}$$

In this case, the projection to S^3 is the disk

$$\tilde{U}_a : (r, \phi) \longmapsto \left(\frac{e^{i\phi}}{\sqrt{1 + e^{-2r}|a|^2}}, \frac{a}{\sqrt{1 + e^{-2r}|a|^2}} \right),$$

which bounds the Hopf fiber $(e^{i\phi}, 0)$. Each disk $\tilde{U}_a$ is clearly transverse to the Hopf flow away from the boundary and the return map is the identity since the flow is periodic. See Figure 9.5 for an illustration.

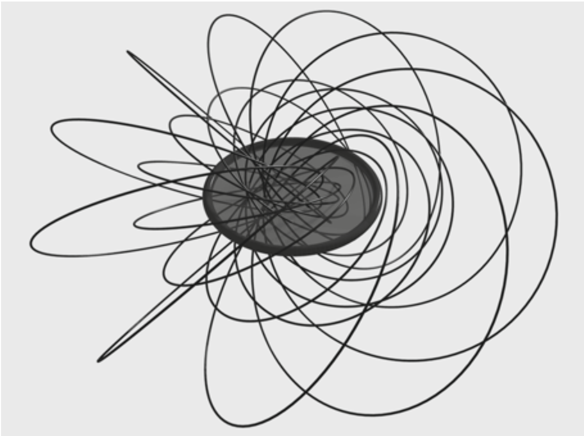


Figure 9.5: The trivial global disk-like surface of section for the Hopf flow after stereographic projection, and some Hopf fibers.

Remark 9.4.2. We point out that this example does *not* satisfy the non-degeneracy condition for its asymptotic orbit, as defined for example in Definition 7.3.1 or the equivalent condition of Equation (7.11). Instead, we have a Morse–Bott condition not only in the direction of the orbit, but also transverse to the orbit: each Hopf fiber can be the asymptote of a similar looking plane. From Formula (9.7), we directly see that we still have exponential convergence to the asymptote; we shall refer to this observation in the proof of Corollary 17.1.4.

9.4.2 General results

In the following we assume that $\Sigma \subset \mathbb{C}^2$ is a star-shaped hypersurface. It follows that the restriction of the one-form

$$\lambda = \frac{1}{2}(x_1 dy_1 - y_1 dx_1 + x_2 dy_2 - y_2 dx_2)$$

to Σ defines a contact form, see Example 2.6.6. The following theorem is due to Hryniewicz [133, 134].

Theorem 9.4.3 (Hryniewicz)**.** *Assume that $\Sigma \subset \mathbb{C}^2$ is a dynamically convex star-shaped hypersurface and $\gamma \in C^\infty(S^1, \Sigma)$ is an unknotted periodic Reeb orbit of period τ whose self-linking number satisfies $s\ell(\gamma) = -1$. Then γ bounds a global surface of section. Moreover, each periodic orbit which corresponds to a fixed point of the Poincaré return map of the global surface of section is unknotted and has self-linking number -1.*

For convenience of the reader, we remind the reader that a star-shaped hypersurface is dynamically convex if the Conley–Zehnder indices of all periodic Reeb orbits are at least 3, see Definition 1.3.1. The somewhat technical definition of Conley–Zehnder index is given in Chapter 10. Intuitively, the Conley–Zehnder index of an orbit is a winding number of the linearized flow.

Remark 9.4.4. The fact that a periodic orbit corresponding to a fixed point of the Poincaré return map itself is unknotted and has self-linking number -1 has an interesting consequence. Namely, if Σ is dynamically convex, we can apply Theorem 9.4.3 again to this orbit to see that it also bounds a global surface of section. This leads to Hryniewicz' theory of systems of global surfaces of section [134].

Remark 9.4.5. Under the assumption that all periodic orbits on the star-shaped hypersurface $\Sigma \subset \mathbb{C}^2$ are non-degenerate, there exists an interesting improvement of the theorem of Hryniewicz which is due to Hryniewicz and Salomão [135]. Namely if one assumes that all periodic orbits of Conley–Zehnder index 2 are linked to γ, then γ still bounds a global surface of section.

The following theorem is due to Hofer, Wysocki, and Zehnder [123].

Theorem 9.4.6 (Hofer–Wysocki–Zehnder)**.** *Assume that $\Sigma \subset \mathbb{C}^2$ is a star-shaped hypersurface, then there exists an unknotted periodic Reeb orbit on Σ with self-linking number -1.*

Remark 9.4.7. Actually Hofer, Wysocki, and Zehnder show in [123] an even stronger result. Namely they prove that every Reeb vector field on S^3 admits an unknotted periodic orbit with self-linking number -1, i.e., the contact form does not necessarily need to come from a starshaped hypersurface in $\mathbb{C}^2$, so that the contact structure does not need to be tight but can also be overtwisted.

Combining the above two theorems we immediately obtain the following corollary.

Corollary 9.4.8 (Hofer–Wysocki–Zehnder)**.** *Assume that $\Sigma \subset \mathbb{C}^2$ is a star-shaped, dynamically convex hypersurface, then Σ admits a global surface of section.*

Remark 9.4.9. Historically, this corollary was first proved by Hofer–Wysocki–Zehnder in the paper [126], before the theorem of Hryniewicz was available. This comes from the fact that both Hryniewicz as well as Hofer, Wysocki, and Zehnder use holomorphic curves to prove their theorems. Therefore Hofer, Wysocki, and Zehnder could use them to prove directly the existence of a global surface of section in their groundbreaking work [126].

Recall from Section 8.3.3 in Chapter 8 the Birkhoff set which consists of pairs (μ, c) of a mass ratio μ and an energy level c below the first critical value which have the property that the retrograde periodic orbit is non-degenerate and linked to every periodic orbit of smaller period. In this monograph we show the following theorem.

Theorem 9.4.10. *Assume that $(\mu, c) \in \mathfrak{B}$. Then the retrograde periodic orbit γ_R bounds a global surface of section on $\Sigma_{\mu,c}$.*

9.5 Invariant global surfaces of section

Assume that $\rho \in \mathrm{Diff}(S^3)$ is a smooth involution on the three-dimensional sphere, i.e., $\rho^2 = \mathrm{id}$ and $X \in \Gamma(TS^3)$ is a non-vanishing vector field on S^3 which is anti-invariant under the involution ρ in the sense that

$$\rho^* X = -X. \tag{9.8}$$

A (disk-like) global surface of section $D \subset S^3$ is called *invariant* if $\rho(D) = D$. By abuse of notation we denote the restriction of ρ to D again by the same letter.

Lemma 9.5.1. *Assume that $D \subset S^3$ is an invariant global surface of section for X. Then the fixed point set $\mathrm{Fix}(\rho) \subset D$ is a simple arc intersecting the boundary ∂D transversely.*

Proof. We first note that ρ is an orientation reversing involution of D. Indeed, this follows from (9.8) in view of the fact that the vector field is tangent to the boundary ∂D. Therefore $\mathrm{Fix}(\rho)$ is a one-dimensional submanifold of D. That it is transverse to ∂D follows again from (9.8). It follows from Brouwer's fixed point theorem, see for example [111, Theorem 1.9.], that $\mathrm{Fix}(\rho)$ is not empty. In particular, it is a finite union of circles and intervals. We claim that there are no circles. To see that we argue by contradiction and assume that the fixed point set of ρ contains a circle. The complement of this circle consists of two connected components, one of them containing the boundary of ∂D. The involution ρ then has to interchange these two connected components. However, the boundary of ∂D is invariant under ρ and this leads to the desired contradiction. Consequently, the fixed points set consists just of a finite union of intervals. It remains to show that there is just one interval. To see that note that the complement of an interval consists again of two connected components which are interchanged by ρ. Therefore there cannot be additional fixed points and the lemma is proved. □

Remark 9.5.2. The lemma above is actually an easy case of a much more general result due to Brouwer [47] and Kérékjartò [147], which says that a topological involution just defined in the interior of the disk is topologically conjugated to a reflection at a line. We refer to [64] for a modern exposition of this result.

We observe in particular that the binding orbit of an invariant global surface of section is necessarily a symmetric periodic orbit. Moreover, if $\psi\colon \mathring{D} \to \mathring{D}$ is the Poincaré return map it follows as in (2.18) that ψ satisfies with ρ the commutation relation

$$\rho\psi\rho = \psi^{-1}. \tag{9.9}$$

Note that the energy hypersurface of the restricted three-body problem after Levi-Civita regularization is invariant under the anti-symplectic involution

$$\rho\colon \mathbb{C}^2 \to \mathbb{C}^2, \quad (z_1, z_2) \mapsto (\overline{z}_1, -\overline{z}_2) \tag{9.10}$$

and the retrograde periodic orbit is symmetric with respect to this involution. We show the following improvement of Theorem 9.4.10.

Theorem 9.5.3. *Assume that $(\mu, c) \in \mathfrak{B}$. Then the retrograde periodic orbit γ_R bounds an invariant global surface of section on $\Sigma_{\mu,c}$.*

Remark 9.5.4. If $\Sigma \subset \mathbb{C}$ is a starshaped, dynamically convex hypersurface satisfying $\rho(\Sigma) = \Sigma$ for the anti-symplectic involution ρ given by (9.10), then Σ admits an invariant global surface of section. In particular, there exists an unknotted, symmetric periodic orbit of self-linking number -1 on Σ. This symmetric analog of Corollary 9.4.8 was proved in [93]. If instead of the anti-symplectic involution ρ, the hypersurface Σ is invariant under the symplectic involution $-\mathrm{id}$, i.e., $\Sigma = -\Sigma$, then it was proved in [136], that on the quotient $\Sigma/\{-\mathrm{id}\} \cong \mathbb{R}P^3$ there exists a rational disk-like global surface of section, namely the disk is only required to be embedded in the interior but doubly covered at the boundary. In particular, there exists an unknotted periodic orbit of self-linking number -1 which doubly covers a non-contractible orbit on the quotient $\Sigma/\{-\mathrm{id}\}$. Note that the symplectic involution $-\mathrm{id}$ and the anti-symplectic involution ρ commute. If Σ is invariant under both involutions, i.e., $\Sigma = -\Sigma = \rho(\Sigma)$, then it seems to be an unexplored question if one can arrange compatibility of the global surface of section with both symmetries. In particular, does there necessarily exist an unknotted symmetric periodic orbit of self-linking number -1? If (μ, c) lies in the Birkhoff set, the retrograde periodic orbit satisfies these requirements.

9.6 Fixed points and periodic points

In his lifelong quest for periodic orbits Poincaré propagated the concept of a global surface of section in [205], because in the presence of a global surface of section the search for periodic orbits is reduced to the search of periodic points of the Poincaré return map. The surface of section that Poincaré used at first in [205], was actually a half-plane rather than an annulus. In his famous last paper [207] Poincaré considered an annulus type global surface of section. This was taken up by Birkhoff in [38] who proved Poincaré's last theorem. This theorem asserts that an area preserving map of an annulus which rotates the two boundary components in opposite directions has at least two fixed points. In [90, 91] Franks proved the following related theorem.

Theorem 9.6.1 (Franks)**.** *Assume an area preserving homeomorphism of an open annulus admits a periodic point. Then it admits infinitely many periodic points.*

On the other hand, Brouwer's translation theorem [46] asserts that

Theorem 9.6.2 (Brouwer)**.** *An area preserving homeomorphism of the open disk admits a fixed point.*

If we restrict an area preserving homeomorphism to the complement of one of its fixed points we obtain an area preserving homeomorphism of the open annulus. This implies the following corollary.

Corollary 9.6.3. *An area preserving homeomorphism of an open disk admits either one or infinitely many periodic points.*

We obtain a further corollary by making two observations. First of all, for a vector field admitting a global surface of section, there is a one-to-one correspondence between periodic orbits different from the binding and periodic points of the Poincaré return map. By Lemma 9.1.4 the return map is area preserving, and so Franks' theorem implies.

Corollary 9.6.4. *Assume that $\omega \in \Omega^2(S^3)$ is a Hamiltonian structure on S^3, $X \in \Gamma(\ker \omega)$ a Hamiltonian vector field whose flow admits a global surface of section. Then X has either two or infinitely many periodic orbits.*

The fixed point guaranteed by Theorem 9.6.2 is of special interest in view of Theorem 9.4.3, because under the assumptions of this theorem the periodic orbit corresponding to the fixed point is itself unknotted and has self-linking number -1 so that it bounds in the dynamically convex case another global surface of section. The amazing thing about Theorem 9.6.2 is that the homeomorphism of the open disk is not required to extend continuously to the boundary of the disk. If it does, then Theorem 9.6.2 just follows from Brouwer's fixed point theorem, see for example [111, Theorem 1.9]. In this case the assumption that the homeomorphism is area preserving is not needed at all. However, if the homeomorphism does not extend continuously to the boundary, then the condition that the map is area preserving is essential. In fact, the open disk is just homeomorphic to the two-dimensional plane and a translation of the plane is an example of a homeomorphism without fixed points. To prove Brouwer's translation theorem one first shows that if an orientation preserving homeomorphism of the open disk has a periodic point it has to have a fixed point. We refer to [82] for a modern account of this remarkable fact. Hence one is left with the case that the homeomorphism has no periodic point at all and one shows in this case that is has to be a translation which then contradicts the assumption that the homeomorphism is area preserving. A modern treatment of this second step together with the precise definition what a translation is can be found in [89], see also [107]. It was observed by Kang in [143] that a quite different argument for this fact can be given if the global surface of section is symmetric. Moreover, in this case one can find a fixed point of the Poincaré return map which corresponds to a symmetric periodic orbit. We discuss this in the next section.

9.7 Reversible maps and symmetric fixed points

Let (D, ω) be a closed two-dimensional disk together with an area form $\omega \in \Omega^2(D)$ and suppose that $\rho \in \mathrm{Diff}(D)$ is an anti-symplectic involution, i.e.,

$$\rho^2 = \mathrm{id}, \quad \rho^* \omega = -\omega.$$

Moreover, suppose that $\psi \in \mathrm{Diff}(\mathring{D})$ is an area preserving diffeomorphism of the interior of the disk, i.e., it holds that

$$\psi^* \omega = \omega,$$

which satisfies with ρ the commutation relation

$$\rho \psi \rho = \psi^{-1}. \tag{9.11}$$

This is the situation one faces by (9.9) if one considers the Poincaré return map of a symmetric global surface of section for a Hamiltonian vector field of a Hamiltonian structure on S^3. A diffeomorphism satisfying (9.11) is called *reversible*. The following result was proved by Kang in [143].

Lemma 9.7.1 (Kang). *Under the above assumptions there exists a common fixed point of ρ and ψ in $\mathring{D}$, i.e., a point $x \in \mathring{D}$ satisfying*

$$\rho(x) = x, \quad \psi(x) = x.$$

Proof. In view of (9.11) we obtain

$$(\psi\rho)^2 = \psi\psi^{-1} = \mathrm{id}$$

so that $\psi\rho$ is again an involution. Moreover, because ρ is anti-symplectic and ψ is symplectic, the composition is also anti-symplectic, so that we have

$$(\psi\rho)^*\omega = -\omega.$$

By Lemma 9.5.1 we know that $\mathrm{Fix}(\rho) \subset D$ is a simple arc intersecting the boundary ∂D transversely. Because ρ is anti-symplectic we conclude that both connected components of the complement of $\mathrm{Fix}(\rho)$ have the same area.

The anti-symplectic involution $\psi\rho$ is only defined in the interior of the disk. However, by the theorem of Brouwer and Kérékjartò mentioned in Remark 9.5.2 the involution $\psi\rho$ is topologically conjugated to the reflection at a line and therefore its fixed point set still consists of an arc whose complement has two connected components. Because $\psi\rho$ is anti-symplectic the two connected components of the complement of its fixed point set have the same area as well.

It follows that $\mathrm{Fix}(\rho)$ and $\mathrm{Fix}(\psi\rho)$ intersect. This means that there exists $x \in \mathring{D}$ with the property that

$$\rho(x) = x, \quad \psi\rho(x) = x$$

or equivalently

$$\rho(x) = x, \quad \psi(x) = x.$$

This finishes the proof of the lemma. $\square$

If $\psi\colon \mathring{D} \to \mathring{D}$ is a reversible map, we call, following [143], a point $x \in \mathring{D}$ a *symmetric periodic point* of ψ if there exist $k, \ell \in \mathbb{N}$ with the property that

$$\psi^k(x) = x, \quad \psi^\ell(x) = \rho(x).$$

If $k = \ell = 1$ then the symmetric periodic point is called a *symmetric fixed point* whose existence was discussed in Lemma 9.7.1. The minimal k for which $\psi^k(x) = x$ holds is referred to as the *period* of x and abbreviated by $k(x)$. Then there exists a unique

$$\ell(x) \in \mathbb{Z}/k(x)\mathbb{Z}$$

such that

$$\psi^{\ell(x)}(x) = \rho(x).$$

Note that the period only depends on the orbit of x, in particular it holds that

$$k(\psi(x)) = k(x).$$

This is not true for $\ell(x)$. In particular, from (9.11) together with the fact that ρ is an involution we obtain the relation

$$\rho = \psi\rho\psi.$$

Using that, we compute

$$\psi^{\ell(x)+1}(x) = \psi\rho(x) = \psi^2\rho\psi(x),$$

implying that

$$\psi^{\ell-2}(\psi(x)) = \rho(\psi(x)).$$

Therefore it holds that

$$\ell(\psi(x)) = \ell(x) - 2 \in \mathbb{Z}/k(x)\mathbb{Z}.$$

If the period is odd then there exists a unique point in the orbit of x which lies on the fixed point set $\mathrm{Fix}(\rho)$. If the period is even, then there are two cases. Either ℓ is also even for every point in the orbit of x and there are precisely two points in the orbit of x which lie on $\mathrm{Fix}(\rho)$ or ℓ is odd for every point in the orbit of x and there is no point in the orbit of x which lies on the fixed point set of ρ.

In [143] Kang proved the following analog of Franks' theorem (Theorem 9.6.1) for reversible maps.

Theorem 9.7.2 (Kang)**.** *Assume an area preserving reversible homeomorphism of an open annulus admits a periodic point, then it admits infinitely many symmetric periodic points.*

Remark 9.7.3. A surprising feature of Kang's theorem is the fact that the periodic point does not need to be symmetric in order to guarantee infinitely many symmetric periodic points.

Combining Lemma 9.7.1 with Theorem 9.7.2 we obtain the following corollary.

Corollary 9.7.4. *A reversible anti-symplectic map of the open disk admits either one or infinitely many symmetric periodic points.*

Because orbits of symmetric periodic points of the Poincaré return map of an invariant global surface of section together with the binding orbit correspond to symmetric periodic orbits we obtain the following corollary.

Corollary 9.7.5. *Assume that $\omega \in \Omega^2(S^3)$ is a Hamiltonian structure on S^3 which is anti-invariant under an involution of S^3 and $X \in \Gamma(\ker\omega)$ is a Hamiltonian vector field whose flow admits an invariant global surface of section. Then X has either two or infinitely many symmetric periodic orbits.*

Chapter 10

The Maslov Index

10.1 The Maslov index for loops

Assume that (V, ω) is a finite-dimensional symplectic vector space. By choosing a symplectic basis $\{e_1, \dots, e_n, f_1, \dots, f_n\}$, i.e., a basis of V satisfying

$$\omega(e_i, e_j) = \omega(f_i, f_j) = 0,\ \omega(e_i, f_j) = \delta_{ij},\ 1 \leq i, j \leq n$$

we can identify V with $\mathbb{C}^n$ by mapping a vector $\xi = \xi^1 + \xi^2 \in V$ with $\xi^1 = \sum_{j=1}^n \xi_j^1 e_j$, $\xi^2 = \sum_{j=1}^n \xi_j^2 f_j$ to the vector $(\xi_1^1 + i\xi_1^2, \dots, \xi_n^1 + i\xi_n^2) \in \mathbb{C}^n$. We then define the *Lagrangian Grassmannian*

$$\Lambda = \Lambda(n),$$

which is the manifold consisting of all Lagrangian subspaces $L \subset \mathbb{C}^n$. For example, since any one-dimensional linear subspace of $\mathbb{C}$ is Lagrangian, we have

$$\Lambda(1) = \mathbb{R}P^1 \cong S^1.$$

To give another example, the Lagrangian Grassmannian $\Lambda(2)$ is diffeomorphic to

$$\Lambda(2) \cong \mathbb{R} \times S^2/(t, x) \sim (t-1, a(x)\,),$$

where a is the antipodal map on S^2. This Lagrangian Grassmannian is not orientable. The explanation in the next paragraph can be used to prove these assertions.

In general, the Lagrangian Grassmannian has the structure of a *homogeneous space*, meaning a manifold with a transitive group action by a finite-dimensional Lie group G; such a space can be written as a quotient G/H where H is a subgroup in G. To see this we first observe that the group $U(n)$ acts on the Lagrangian Grassmannian

$$U(n) \times \Lambda(n) \to \Lambda(n), \quad (U, L) \mapsto UL.$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_10

Indeed, UL is Lagrangian, because U acts as a symplectic matrix. That this action is transitive can be seen in the following way. On a given Lagrangian $L \subset \mathbb{C}^n$ choose an orthonormal basis $e'_1, \ldots, e'_n$ of L, where orthonormal refers of course to the standard inner product on $\mathbb{C}^n$. Putting $f'_j = ie_j$ for $1 \le j \le n$ gives rise to a symplectic, orthonormal basis $\{e'_1, \ldots, e'_n, f'_1, \ldots, f'_n\}$ of $\mathbb{C}^n$. Now define $U\colon \mathbb{C}^n \to \mathbb{C}^n$ as the linear map which maps e_j to e'_j and f_j to f'_j. This proves transitivity. The stabilizer of the $U(n)$ action on $\Lambda(n)$ can be identified with the group $O(n)$, namely the ambiguity in choosing an orthonormal basis on the Lagrangian subspace. Therefore, the Lagrangian Grassmannian becomes the homogeneous space

$$\Lambda(n) = U(n)/O(n).$$

From here, we can quickly deduce the two examples given earlier. Following [16] we next discuss the fundamental group of the Lagrangian Grassmannian. Consider the map

$$\rho\colon U(n)/O(n) \to S^1, \quad [A] \mapsto \det A^2.$$

Note that because the determinant of a matrix in $O(n)$ is plus or minus one, the map ρ is well defined, independent of the choice of the representative $A \in U(n)$. This gives rise to a fiber bundle

$$\begin{tikzcd} SU(n)/SO(n) \arrow[r] & U(n)/O(n) \arrow[d, "\rho"] \\ & S^1. \end{tikzcd}$$

Since the fiber $SU(n)/SO(n)$ is simply connected the long exact homotopy sequence tells us that the induced homomorphism

$$\rho_*\colon \pi_1\big(U(n)/O(n)\big) \to \pi_1(S^1)$$

is an isomorphism. We state this fact as a theorem.

Theorem 10.1.1. *The fundamental group of the Lagrangian Grassmannian satisfies*

$$\pi_1(\Lambda) \cong \mathbb{Z}.$$

Moreover, an explicit isomorphism is given by the map $\rho_*\colon \pi_1(\Lambda) \to \pi_1(S^1)$.

If $\lambda\colon S^1 \to \Lambda$ is a continuous loop of Lagrangian subspaces we obtain a continuous map

$$\rho \circ \lambda\colon S^1 \to S^1$$

and we define the *Maslov index of a loop* as

$$\mu(\lambda) := \deg(\rho \circ \lambda) \in \mathbb{Z}.$$

In view of Theorem 10.1.1 we can alternatively characterize the Maslov index as

$$\mu(\lambda) = [\lambda] \in \pi_1(\Lambda) = \mathbb{Z}.$$

It is not clear from this definition how to generalize the definition of the Maslov index from a loop of Lagrangian subspaces to a *path* of Lagrangian subspaces $\lambda\colon [0,1] \to \Lambda$. In order to find such a generalization, which is needed to define the Conley–Zehnder index, we discuss Arnold's characterization of the Maslov index as an intersection number with the *Maslov* (*pseudoco*)*cycle* [16]. We refer to the book of McDuff and Salamon, [173], for a definition of this notion, but intuitively this is the image of a manifold with "boundary" that has codimension at least 2.

10.2 The Maslov cycle

In order to define the Maslov pseudo-cocycle we fix a basepoint $L_0 \in \Lambda(n)$ and define for $0 \leq k \leq n$

$$\Lambda^k = \Lambda^k_{L_0}(n) = \{L \in \Lambda(n) : \dim(L \cap L_0) = k\}.$$

For each $0 \leq k \leq n$ the space Λ^k is a submanifold of Λ and the whole Lagrangian Grassmannian is stratified as

$$\Lambda = \bigcup_{k=0}^{n} \Lambda^k. \tag{10.1}$$

We need the following proposition

Proposition 10.2.1. *For $0 \leq k \leq n$ the codimension of $\Lambda^k \subset \Lambda$ is given by*

$$\operatorname{codim}(\Lambda^k, \Lambda) = \frac{k(k+1)}{2}.$$

Proof. We first compute the dimension of the Lagrangian Grassmannian. Because $\Lambda(n) = U(n)/O(n)$ we obtain

$$\begin{aligned} \dim \Lambda(n) &= \dim U(n)/O(n) = \dim U(n) - \dim O(n) \\ &= n^2 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2}. \end{aligned} \tag{10.2}$$

Recall further that the usual Grassmannian

$$G(n,k) = \{V \subset \mathbb{R}^n : \dim V = k\}$$

of k-planes in $\mathbb{R}^n$ can be interpreted as the homogeneous space

$$G(n,k) = O(n)/O(k) \times O(n-k)$$

and therefore its dimension is given by

$$\begin{aligned}\dim G(n,k) &= \dim O(n) - \dim O(k) - \dim O(n-k) &(10.3)\\ &= \frac{n(n-1)}{2} - \frac{k(k-1)}{2} - \frac{(n-k)(n-k-1)}{2} \\ &= k(n-k). &(10.4)\end{aligned}$$

Given a Lagrangian $L \in \Lambda^k_{L_0}(n)$, denote by $(L \cap L_0)^\perp$ the orthogonal complement of $L \cap L_0$ in L_0. We obtain a symplectic splitting

$$\mathbb{C}^n = \Big(L \cap L_0 \oplus i(L \cap L_0)\Big) \oplus \Big((L \cap L_0)^\perp \oplus i(L \cap L_0)^\perp\Big) =: V_0 \oplus V_1.$$

Note that V_0 and V_1 are symplectic subvector spaces of $\mathbb{C}^n$ of dimension $\dim V_0 = 2k$ and $\dim V_1 = 2(n-k)$. Moreover, $(L \cap L_0)^\perp \subset V_1$ as well as $L \cap V_1$ are Lagrangian subspaces of V_1 which satisfy $(L \cap L_0)^\perp \cap (L \cap V_1) = \{0\}$. Therefore we can interpret

$$L \cap V_1 \in \Lambda^0_{(L \cap L_0)^\perp}(n-k).$$

Furthermore, $L \cap V_0 = L \cap L_0$ is a k-dimensional subspace in L and therefore the projection

$$\Lambda^k_{L_0}(n) \to G(n,k), \quad L \mapsto L \cap L_0$$

gives rise to a fiber bundle

$$\begin{tikzcd} \Lambda^0(n-k) \arrow[r] & \Lambda^k(n) \arrow[d] \\ & G(n,k). \end{tikzcd}$$

Since $\Lambda^0(n-k)$ is an open subset of $\Lambda(n-k)$, its dimension equals by (10.2)

$$\dim \Lambda^0(n-k) = \frac{(n-k)(n-k+1)}{2}.$$

Using (10.3) we obtain

$$\begin{aligned}\dim \Lambda^k(n) &= \dim \Lambda^0(n-k) + \dim G(n,k) &(10.5)\\ &= \frac{(n-k)(n-k+1)}{2} + k(n-k) = \frac{n(n+1)}{2} - \frac{k(k+1)}{2}.\end{aligned}$$

Combining (10.2) and (10.5) the proposition follows. $\square$

Note that if we choose $k=1$ in the above proposition we obtain that $\Lambda^1 \subset \Lambda$ is a codimension 1 submanifold of Λ. However, Λ^1 is in general non-compact, and its closure is given by

$$\overline{\Lambda^1} = \bigcup_{k=1}^n \Lambda^k.$$

In particular, the boundary of Λ^1 is given by

$$\overline{\Lambda^1} \setminus \Lambda^1 = \bigcup_{k=2}^{n} \Lambda^k$$

and from Proposition 10.2.1 we deduce that

$$\dim \Lambda^1 - \dim \Lambda^k \geq 2$$

which means that the boundary of Λ^1 has dimension at least 2 less than Λ^1. This property is crucial for intersection theoretic arguments with Λ^1.

Because the boundary of Λ^1 has dimension at least 2 it follows that if Λ^1 is in addition orientable, then it meets the requirements of a pseudocycle as in [173]. Results of Kahn, Schwarz, and Zinger [141, 217, 245], then show how to associate a homology class to a pseudocycle. Unfortunately, Λ^1 is not always orientable. This is related to a theorem of D. Fuks [95] which tells us that the Lagrangian Grassmannian $\Lambda(n)$ is orientable if and only if n is odd. As was noted by Arnold [16] Λ^1 is always coorientable. Hence if n is odd, then the submanifold $\Lambda^1(n)$ is a pseudocycle in the sense of McDuff and Salamon. Fortunately, the intersection number of Λ^1 with a loop of Lagrangian subspaces can always be defined independently of these theoretic considerations due to the fact that Λ^1 can be cooriented. A cartoon is sketched in Figure 10.1. Our next goal is to show how this coorientation works.

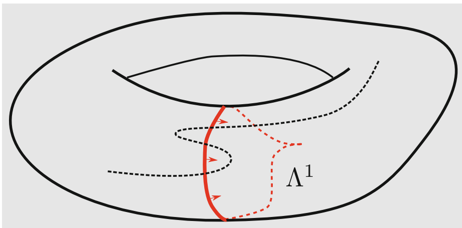


Figure 10.1: A path of Lagrangians intersecting the Maslov cycle.

We first describe some natural charts for the Lagrangian Grassmannian. Given $L_1 \in \Lambda$ choose a Lagrangian complement L_2 of L_1, i.e., $L_2 \in \Lambda$ such that $L_2 \cap L_1 = \{0\}$. It follows that $L_1 \in \Lambda^0_{L_2}$. We now explain how to construct a vector space structure on $\Lambda^0_{L_2}$ for which L_1 becomes the origin. This is done by identifying $\Lambda^0_{L_2}$ with $S^2(L_1)$, the space of quadratic forms on L_1. Namely, given $L \in \Lambda^0_{L_2}$ for each $v \in L_1$ there exists a unique $w_v \in L_2$ such that $v + w_v \in L$. We define

$$\Lambda^0_{L_2} \to S^2(L_1), \quad L \mapsto Q_L = Q_L^{L_1,L_2} \tag{10.6}$$

where

$$Q_L(v) = \omega(v, w_v), \quad v \in L_1.$$

We describe this procedure in coordinates. Namely, we choose a basis $\{e_1, \dots, e_n\}$ of L_1. Then there exists a unique basis $\{f_1, \dots, f_n\}$ of the Lagrangian complement L_2 of L_1 such that $\{e_1, \dots, e_n, f_1, \dots, f_n\}$ is a symplectic basis of $\mathbb{C}^n$. Using these bases we identify $L_1 = \mathbb{R}^n \subset \mathbb{C}^n$ and $L_2 = i\mathbb{R}^n \subset \mathbb{C}^n$. Now if $v = v_1 + iv_2, w = w_1 + iw_2 \in \mathbb{C}^n$ with $v_1, v_2, w_1, w_2 \in \mathbb{R}^n$, then the symplectic form is given by

$$\omega(v, w) = \langle v_1, w_2 \rangle - \langle v_2, w_1 \rangle$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^n$. Now if $L \subset \mathbb{C}^n$ is a Lagrangian satisfying $L \cap L_2 = \{0\}$ we can write L as

$$L = \{x + iSx : x \in \mathbb{R}^n\} =: \Gamma_S$$

namely as the graph of a linear map $S \colon \mathbb{R}^n \to \mathbb{R}^n$. The fact that L is Lagrangian translates into the fact that for every $x, y \in \mathbb{R}^n$ we have

$$0 = \omega(x + iSx, y + iSy) = \langle x, Sy \rangle - \langle y, Sx \rangle$$

which is equivalent to the assertion that

$$\langle x, Sy \rangle = \langle y, Sx \rangle, \quad x, y \in \mathbb{R}^n$$

meaning that S is symmetric. Now for $x \in \mathbb{R}^n$ we compute

$$Q_L(x) = \omega(x, iSx) = \langle x, Sx \rangle.$$

Remark 10.2.2. The natural charts lead to a new proof of the dimension formula for the Lagrangian Grassmannian. Indeed,

$$\dim S^2(L_1) = \frac{n(n+1)}{2}$$

in accordance with (10.2).

We next show how the natural charts lead to an identification of the tangent space $T_{L_1}\Lambda$ with the space of quadratic forms $S^2(L_1)$ on L_1. Indeed, given a Lagrangian complement L_2 of L_1, i.e., an element $L_2 \in \Lambda^0_{L_1}$, the natural chart obtained from L_2 leads to a vector space isomorphism

$$\Phi_{L_2} \colon T_{L_1}\Lambda \to S^2(L_1), \quad L \mapsto Q_L^{L_1, L_2}.$$

Our next lemma shows that this vector space isomorphism does not depend on the chart chosen.

Lemma 10.2.3. *The vector space isomorphism Φ_{L_2} is independent of L_2, i.e., we have a canonical identification of $T_{L_1}\Lambda$ with $S^2(L_1)$.*

Proof. We can assume without loss of generality that $L_1 = \mathbb{R}^n \subset \mathbb{C}^n$. Then an arbitrary Lagrangian complement of L_1 is given by

$$L_2 = \{By + iy : y \in \mathbb{R}^n\}$$

where B is a real symmetric $n \times n$ matrix. Now consider a smooth path $L\colon (-\epsilon, \epsilon) \to \Lambda$ satisfying $L(0) = L_1$ and such that

$$L(t) = \{x + iA(t)x : x \in \mathbb{R}^n\}$$

where $A(t)$ is a symmetric matrix such that $A(0) = 0$. For $x \in \mathbb{R}^n = L_1$ we next compute $Q^{L_1,L_2}_{L(t)}(x)$. For that purpose let $w_x(t)$ be the unique vector in L_2 such that

$$x + w_x(t) \in L(t).$$

Since $w_x(t) \in L_2$ there exists a unique $y(t) \in \mathbb{R}^n$ such that

$$w_x(t) = By(t) + iy(t).$$

We therefore obtain

$$x + w_x(t) = x + By(t) + iy(t) \in L(t)$$

which implies that

$$y(t) = A(t)(x + By(t)). \tag{10.7}$$

Moreover,

$$Q^{L_1,L_2}_{L(t)}(x) = \omega(x, w_x(t)) = \omega(x, By(t) + iy(t)) = \langle x, y(t)\rangle$$

such that

$$\left.\frac{d}{dt}\right|_{t=0} Q^{L_1,L_2}_{L(t)}(x) = \langle x, y'(0)\rangle. \tag{10.8}$$

Since $L(0) = L_1$ it follows that $w_x(0) = 0$ and therefore $y(0) = 0$. Using this in combination with $A(0) = 0$, we obtain by differentiating (10.7)

$$y'(0) = A'(0)(x + By(0)) + A(0)(x + By'(0)) = A'(0)x.$$

Plugging this into (10.8) we get

$$\left.\frac{d}{dt}\right|_{t=0} Q^{L_1,L_2}_{L(t)}(x) = \langle x, A'(0)x\rangle. \tag{10.9}$$

This is independent of B and therefore $\left.\frac{d}{dt}\right|_{t=0} Q^{L_1,L_2}_{L(t)}(x)$ does not depend on the choice of the Lagrangian complement L_2. This proves the lemma. $\square$

In view of the above lemma we denote for $L \in \Lambda$ and $\widehat{L} \in T_L\Lambda$ the uniquely determined quadratic form on L by

$$Q^{\widehat{L}} \in S^2(L).$$

We can use this form to characterize the tangent space of $\Lambda^k_{L_0}$.

Lemma 10.2.4. *Assume that $L_0 \in \Lambda$, $k \in \{0, \dots, n\}$, and $L \in \Lambda^k_{L_0}$, then*

$$T_L\Lambda^k_{L_0} = \left\{\widehat{L} \in T_L\Lambda : Q^{\widehat{L}}\big|_{L_0 \cap L} = 0\right\}.$$

Proof. We decompose $\mathbb{C}^n$ as

$$\mathbb{C}^n = \mathbb{R}^k \times \mathbb{R}^{n-k} \times \mathbb{R}^k \times \mathbb{R}^{n-k}.$$

We can assume without loss of generality that

$$L = \mathbb{R}^k \times \mathbb{R}^{n-k} \times \{0\} \times \{0\} = \mathbb{R}^n \times \{0\}$$

and

$$L_0 = \mathbb{R}^k \times \{0\} \times \{0\} \times \mathbb{R}^{n-k}.$$

Suppose that $A\colon \mathbb{R}^n \to \mathbb{R}^n$ is a matrix satisfying $A = A^T$. We decompose

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where $A_{11}\colon \mathbb{R}^k \to \mathbb{R}^k$, $A_{12}\colon \mathbb{R}^{n-k} \to \mathbb{R}^k$, $A_{21}\colon \mathbb{R}^k \to \mathbb{R}^{n-k}$, $A_{22}\colon \mathbb{R}^{n-k} \to \mathbb{R}^{n-k}$ satisfy

$$A_{11} = A_{11}^T, \quad A_{22} = A_{22}^T, \quad A_{12} = A_{21}^T.$$

The graph of A then can be written as

$$\Gamma_A = \{(x, y, A_{11}x + A_{12}y, A_{21}x + A_{22}y) : x \in \mathbb{R}^k,\ y \in \mathbb{R}^{n-k}\}.$$

Now suppose that $(x_1, y_1, x_2, y_2) \in \Gamma_A \cap L_0$. Since $(x_1, y_1, x_2, y_2) \in \Gamma_A$ we obtain $x_2 = A_{11}x_1 + A_{12}y_1$ and $y_2 = A_{21}x_1 + A_{22}y_1$. Because $(x_1, y_1, x_2, y_2) \in L_0$ we must have $x_2 = y_1 = 0$ and therefore $A_{11}x_1 = 0$ and $y_2 = A_{21}x_1$. We have proved that

$$\Gamma_A \cap L_0 = \{(x_1, 0, 0, A_{21}x_1) : A_{11}x_1 = 0\}. \tag{10.10}$$

If $\widehat{L} \in T_L\Lambda$ we can write $\widehat{L}$ as

$$\widehat{L} = \frac{d}{dt}\bigg|_{t=0} \Gamma_{tA} : \quad A\colon \mathbb{R}^n \to \mathbb{R}^n,\ A = A^T.$$

In view of (10.9) the associated quadratic form is given by

$$Q^{\widehat{L}}\colon L \times L \to \mathbb{R}, \quad z \mapsto \langle z, Az\rangle,\ z = (x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} = \mathbb{R}^n.$$

Its restriction to $L \cap L_0$ becomes

$$Q^{\widehat{L}}\big|_{L\cap L_0} \colon L \cap L_0 \times L \cap L_0 \to \mathbb{R}, \quad x \mapsto \langle x, A_{11} x\rangle,\ x \in \mathbb{R}^k. \tag{10.11}$$

From (10.10) and (10.11) we deduce that the dimension of $\Gamma_{tA} \cap L_0 = k$ for every $t \in \mathbb{R}$ if and only if $Q^{\widehat{L}}\big|_{L\cap L_0} = 0$. This implies that

$$T_L\Lambda^k_{L_0} \subset \Big\{\widehat{L} \in T_L\Lambda : Q^{\widehat{L}}\big|_{L_0\cap L} = 0\Big\}.$$

On the other hand we know from (10.5) that

$$\dim T_L\Lambda^k_{L_0} = \frac{n(n+1)}{2} - \frac{k(k+1)}{2} = \dim\Big\{\widehat{L} \in T_L\Lambda : Q^{\widehat{L}}\big|_{L_0\cap L} = 0\Big\}$$

and therefore

$$T_L\Lambda^k_{L_0} = \Big\{\widehat{L} \in T_L\Lambda : Q^{\widehat{L}}\big|_{L_0\cap L} = 0\Big\}.$$

This finishes the proof of the lemma. □

In view of the previous lemma we can now define the *coorientation* of $\Lambda^1 = \Lambda^1_{L_0}$ in Λ as follows. If $L \in \Lambda^1$ we define

$$[\widehat{L}] \in T_L\Lambda/T_L\Lambda^1 \text{ positive} \iff Q^{\widehat{L}}\big|_{L_0\cap L} \text{ positive.}$$

Indeed, Lemma 10.2.4 tells us that this definition is well defined, independent of the choice of $\widehat{L}$ in its equivalence class in $T_L\Lambda/T_L\Lambda^1$.

We now use the coorientation of Λ^1 to give an alternative characterization of the Maslov index via intersection theory. We fix $L_0 \in \Lambda$ and define for a loop $\lambda\colon S^1 \to \Lambda$ the Maslov index as intersection number between λ and $\Lambda^1 = \Lambda^1_{L_0}$. In order to do that we need the following definition.

Definition 10.2.5. Assume that $\lambda\colon S^1 \to \Lambda$ is a smooth map. We say that λ intersects Λ^1 *transversely* if and only if the following two conditions are satisfied.

(i) For every $t \in S^1$ such that $\lambda(t) \in \Lambda^1$ we have

$$\operatorname{im}(d\lambda(t)) + T_{\lambda(t)}\Lambda^1 = T_{\lambda(t)}\Lambda.$$

(ii) For every $k \geq 2$ it holds that

$$\operatorname{im}\lambda \cap \Lambda^k = \emptyset.$$

One writes $\lambda \pitchfork \Lambda^1$ if λ intersects Λ^1 transversely. It follows from Sard's theorem [177, Chapter 2] and the fact that $\operatorname{codim}(\Lambda^k, \Lambda) \geq 3$ for $k \geq 2$ that after small perturbation of λ we can assume that $\lambda \pitchfork \Lambda^1$. In fact, for this step $\operatorname{codim}(\Lambda^k, \Lambda) \geq 2$ for $k \geq 2$ would already be sufficient. However, we soon need $\operatorname{codim}(\Lambda^k, \Lambda) \geq 3$ for $k \geq 2$ to show invariance of the intersection number under homotopies.

In the following let us assume that $\lambda \pitchfork \Lambda^1$. It follows that $\lambda^{-1}(\Lambda^1)$ is a finite set. For $t \in \lambda^{-1}(\Lambda^1)$ we use the coorientation of Λ^1 to define

$$\nu(t) := \begin{cases} 1 & \partial_t \lambda(t) \in T_{\lambda(t)}\Lambda / T_{\lambda(t)}\Lambda^1 \text{ positive} \\ -1 & \text{else.} \end{cases}$$

We define now the intersection number of λ with Λ^1 as

$$\widetilde{\mu}(\lambda) := \sum_{t \in \lambda^{-1}(\Lambda^1)} \nu(t).$$

Here we understand that if $t \notin \lambda^{-1}(\Lambda^1)$, then $\nu(t) = 0$, so that only finitely many summands in the above sum are different from zero.

Theorem 10.2.6. *The intersection number with the Maslov cycle coincides with the Maslov index, i.e.,* $\widetilde{\mu} = \mu$.

Proof. We prove the theorem in two steps.

Step 1: $\widetilde{\mu}(\lambda)$ only depends on the homotopy class of λ.

In order to prove Step 1, assume that $\lambda_0, \lambda_1 \pitchfork \Lambda^1$ are two loops of Lagrangian subspaces which are homotopic to each other, i.e., we can choose a smooth map

$$\lambda \colon S^1 \times [0,1] \to \Lambda$$

such that

$$\lambda(\cdot, 0) = \lambda_0, \quad \lambda(\cdot, 1) = \lambda_1.$$

Again taking advantage of Sard's theorem and the fact that by Lemma 10.2.1 $\operatorname{codim}(\Lambda^k, \Lambda) \geq 3$ for every $k \geq 2$ we can assume, maybe after perturbing the homotopy λ, that $\lambda \pitchfork \Lambda^1$ in the sense that the following two conditions are met.

(i) For every $(t,r) \in S^1 \times [0,1]$ such that $\lambda(t,r) \in \Lambda^1$ it holds that

$$\operatorname{im}(d\lambda(t,r)) + T_{\lambda(t,r)}\Lambda^1 = T_{\lambda(t,r)}\Lambda.$$

(ii) For every $k \geq 2$ we have

$$\operatorname{im}\lambda \cap \Lambda^k = \emptyset.$$

It now follows from the implicit function theorem that $\lambda^{-1}(\Lambda^1) \subset S^1 \times [0,1]$ is a one-dimensional manifold with boundary. The boundary is given by

$$\partial\Big(\lambda^{-1}(\Lambda^1)\Big) = \Big(\lambda_0^{-1}(\Lambda^1) \times \{0\}\Big) \cup \Big(\lambda_1^{-1}(\Lambda^1) \times \{1\}\Big).$$

The manifold $\lambda^{-1}(\Lambda^1)$ can be oriented as follows. We orient the cylinder by declaring the basis $\{\partial_r, \partial_t\}$ to be positive at every point $(t,r) \in S^1 \times [0,1]$. Suppose that $(t,r) \in \lambda^{-1}(\Lambda^1)$ and $v \neq 0 \in T_{(t,r)}\lambda^{-1}(\Lambda^1)$. Choose $w \in T_{(t,r)}(S^1 \times [0,1])$ such that $\{v, w\}$ is a positive basis of $T_{(t,r)}(S^1 \times [0,1])$. We now declare v to be a positive basis of the one-dimensional vector space $T_{(t,r)}\lambda^{-1}(\Lambda^1)$ if and only if $[d\lambda(w)] \in T\Lambda/T\Lambda^1$ is positive. Let us check that this notion is well defined, independent of the choice of w. Namely, if w' is another choice it follows that $w' = aw + bv$, where $a > 0$ and $b \in \mathbb{R}$. Since $d\lambda(v) \in T\Lambda^1$, we obtain $[d\lambda(w')] = a[d\lambda(w)]$, which due to the positivity of a is positive if and only if $[d\lambda(w)]$ is positive.

Since $\lambda \pitchfork \Lambda^1$ we have $\lambda^{-1}(\Lambda^1) = \lambda^{-1}(\overline{\Lambda^1})$ and therefore $\lambda^{-1}(\Lambda)$ is compact. It follows, see [177, Appendix] that a compact one-dimensional manifold with boundary is a finite union of circles and intervals. For a compact submanifold of the cylinder $S^1 \times [0,1]$ there are three types of intervals. Either both boundary points lie in the first boundary component $S^1 \times \{0\}$, or both boundary points lie in the second component $S^1 \times \{1\}$, or finally one boundary point lies in $S^1 \times \{0\}$ whereas the other boundary point lies in $S^1 \times \{1\}$. In the first case both boundary points contribute with opposite signs $\widetilde{\mu}(\lambda_0)$, in the second case they contribute with opposite signs to $\widetilde{\mu}(\lambda_1)$ and in the third case they contribute with the same sign to $\widetilde{\mu}(\lambda_0)$ and $\widetilde{\mu}(\lambda_1)$. This proves that $\widetilde{\mu}(\lambda_0) = \widetilde{\mu}(\lambda_1)$ and finishes the proof of the first step.

Step 2: We prove the theorem.

As a consequence of Step 1 the intersection number $\widetilde{\mu}$ induces a homomorphism from $\pi_1(\Lambda)$ to $\mathbb{Z}$. By Theorem 10.1.1 we know that $\pi_1(\Lambda) = \mathbb{Z}$. Hence it suffices to show that $\widetilde{\mu}$ agrees with μ on a generator of $\pi_1(\Lambda) = \mathbb{Z}$. Such a generator is given by

$$\lambda(t) = (e^{i\pi t}, 1, \dots, 1)\mathbb{R}^n \subset \mathbb{C}^n, \quad t \in [0,1].$$

For this path we have

$$\rho \circ \lambda(t) = \det \begin{pmatrix} e^{i\pi t} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}^2 = e^{2\pi i t}$$

so that we obtain for the Maslov index

$$\mu(\lambda) = \deg(t \mapsto e^{2\pi i t}) = 1.$$

It remains to compute $\widetilde{\mu}(\lambda)$. A priori the computation of $\widetilde{\mu}(\lambda)$ depends on the choice of the basepoint $L_0 \in \Lambda$ used to define the Maslov cycle. However, the Lagrangian Grassmannian is connected so that between any two basepoints we can always find a smooth path such that a homotopy argument analogous to the one in Step 1 shows that $\widetilde{\mu}(\lambda)$ is independent of the choice of L_0. Taking advantage of this freedom we choose

$$L_0 = (1, i, \dots, i)\mathbb{R}^n.$$

It follows that $\lambda \pitchfork \Lambda^1_{L_0}$ and

$$\lambda^{-1}(\Lambda^1) = \{0\}.$$

It remains to compute the sign at $t = 0$. As Lagrangian complement of $\lambda(0) = \mathbb{R}^n$ we choose $i\mathbb{R}^n$. For $t \in (-\frac{1}{2}, \frac{1}{2})$ we can write the path $\lambda(t)$ as graph

$$\lambda(t) = \{x + iA(t)x : x \in \mathbb{R}^n\}$$

where

$$A(t) = \begin{pmatrix} \tan \pi t & 0 & & 0 \\ 0 & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

By (10.9) the quadratic form $Q^{\lambda'(0)} \in S^2(\mathbb{R}^n)$ is given by

$$Q^{\lambda'(0)}(x) = \langle x, A'(0)x \rangle, \quad x \in \mathbb{R}^n.$$

Now the derivative of the matrix A computes to be

$$A'(t) = \begin{pmatrix} \frac{\pi}{\cos^2 \pi t} & 0 & & 0 \\ 0 & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}$$

and therefore

$$A'(0) = \begin{pmatrix} \pi & 0 & & 0 \\ 0 & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

Moreover,

$$L_0 \cap \lambda(0) = \operatorname{span}\{(1, 0, \ldots, 0)\}$$

and therefore the map

$$Q^{\lambda'(0)}\big|_{L_0 \cap \lambda(0)} : x \mapsto \pi x^2$$

is positive. We conclude that

$$\nu(0) = 1$$

and therefore

$$\widetilde{\mu}(\lambda) = 1 = \mu(\lambda).$$

This finishes the proof of the theorem. □

10.3 The Maslov index for paths

One of the advantages of the intersection theoretic interpretation of the Maslov index is that it generalizes to paths of Lagrangians. This is necessary to define the Conley–Zehnder index. We next explain the definition of the Maslov index for paths due to Robbin and Salamon [211].

Suppose that $\lambda\colon [0,1] \to \Lambda$ is a smooth path of Lagrangian subspaces. Fix a basepoint $L \in \Lambda$. For $t \in [0,1]$ the *crossing form*

$$C(\lambda, L, t) := Q^{\lambda'(t)}\big|_{\lambda(t)\cap L} \tag{10.12}$$

is a quadratic form on the vector space $\lambda(t) \cap L$. An intersection point $t \in \lambda^{-1}(\overline{\Lambda_L^1}) = \lambda^{-1}(\bigcup_{k=1}^n \Lambda_L^k)$ is called a *regular crossing* if and only if the crossing form $C(\lambda, L, t)$ is nonsingular. After a perturbation with fixed endpoints we can assume that the path $\lambda\colon [0,1] \to \Lambda$ has only regular crossings. In fact, we can even assume that for all $t \in (0,1)$ it holds that $\lambda(t) \notin \Lambda_L^k$ for $k \geq 2$. However, $\lambda(0)$ or $\lambda(1)$ might lie in Λ_L^k for some $k \geq 2$.

Suppose now that $\lambda : [0,1] \to \Lambda$ is a path with only regular crossings. Then its Maslov index with respect to the chosen basepoint $L \in \Lambda$ is defined by Robbin and Salamon [211] as

$$\mu_L(\lambda) := \frac{1}{2}\mathrm{sign}C(\lambda, L, 0) + \sum_{0<t<1} \mathrm{sign}C(\lambda, L, t) + \frac{1}{2}\mathrm{sign}C(\lambda, L, 1) \in \frac{1}{2}\mathbb{Z}. \tag{10.13}$$

Here sign refers to the signature of a quadratic form. The Maslov index for paths has the following properties.

Invariance. If $\lambda_0, \lambda_1\colon [0,1] \to \Lambda$ are homotopic to each other with fixed endpoints, then

$$\mu_L(\lambda_0) = \mu_L(\lambda_1).$$

Concatenation. Suppose that $\lambda_0, \lambda_1\colon [0,1] \to \Lambda$ satisfy $\lambda_0(1) = \lambda_1(0)$, then

$$\mu_L(\lambda_0 \# \lambda_1) = \mu_L(\lambda_0) + \mu_L(\lambda_1)$$

where $\lambda_0\#\lambda_1$ refers to the concatenation of the two paths.

Loop. If $\lambda\colon [0,1] \to \Lambda$ is a loop, i.e., $\lambda(0) = \lambda(1)$, then

$$\mu_L(\lambda) = \mu(\lambda)$$

the Maslov index for loops defined before. In particular, for loops the Maslov index does not depend on the choice of the base point $L \in \Lambda$.

Remark 10.3.1. In general the Maslov index for paths depends on the choice of the base point $L \in \Lambda$. However, if $L_0, L_1 \in \Lambda$ and $\lambda\colon [0,1] \to \Lambda$, then the three properties of the Maslov index just described imply that the difference $\mu_{L_0}(\lambda) - \mu_{L_1}(\lambda)$

only depends on $L_0, L_1, \lambda(0), \lambda(1)$, i.e., only the endpoints of the Lagrangian path λ matter. Such indices, which associate to a collection of four Lagrangian subspaces of a symplectic vector space a number, are also known in the literature as Maslov indices and are for example studied in the work by Hörmander [119] or Kashiwara [164].

Remark 10.3.2. There are other ways how to associate to a path of Lagrangian subspaces a Maslov index. We mention here the work of Duistermaat [75]. The Maslov index of Duistermaat has the property that it is independent of the choice of the base point $L \in \Lambda$ at the expense of the concatenation property. The Maslov index of Duistermaat is related to the Maslov index of Robbin and Salamon by a correction term involving a Hörmander–Kashiwara Maslov index.

10.4 The Conley–Zehnder index

We next explain how to define the *Conley–Zehnder index* [62, 63] as a Maslov index. There is a more direct approach to the Conley–Zehnder index, which consists of studying the intersections of paths of symplectic matrices with the *Maslov cycle* in $\mathrm{Sp}(n)$, which is given by

$$V_n = \{A \in \mathrm{Sp}(n) \mid \det(A - \mathrm{id}) = 0\}.$$

This approach is taken in Long's book, [168].

Let us now start with the description of the Conley–Zehnder index as a Maslov index. If (V, ω) is a symplectic vector space the symplectic group $\mathrm{Sp}(V)$ consists of all linear maps $A\colon V \to V$ satisfying $A^*\omega = \omega$. Moreover, if (V, ω) is a symplectic vector space $(V \oplus V, -\omega \oplus \omega)$ is a symplectic vector space as well and has the property that for every $A \in \mathrm{Sp}(V)$ the graph of A

$$\Gamma_A = \{(x, Ax) : x \in V\} \subset V \oplus V$$

is a Lagrangian subspace of $(V \oplus V, -\omega \oplus \omega)$. Indeed, if $(x, Ax), (y, Ay) \in \Gamma_A$ we have

$$(-\omega \oplus \omega)((x, Ax), (y, Ay)) = -\omega(x, y) + \omega(Ax, Ay) = -\omega(x, y) + \omega(x, y) = 0$$

where in the second equality we have used that A is symplectic. In particular, the diagonal

$$\Delta = \Gamma_{\mathrm{id}} = \{(x, x) : x \in V\}$$

is a Lagrangian subspace of $(V \oplus V, -\omega \oplus \omega)$. Suppose now that we are given a smooth path $\Psi\colon [0, 1] \to \mathrm{Sp}(V)$, i.e., a smooth path of linear symplectic maps. We associate to such a path a path of Lagrangian subspaces in $V \oplus V$ by

$$\Gamma_\Psi\colon [0, 1] \to \Lambda(V \oplus V), \quad t \mapsto \Gamma_{\Psi(t)}.$$

We say that a smooth path of symplectic linear maps $\Psi\colon [0,1] \to \mathrm{Sp}(V)$ starting at the identity $\Psi(0) = \mathrm{id}$ is *non-degenerate* if

$$\det(\Psi(1) - \mathrm{id}) \neq 0,$$

i.e., 1 is not an eigenvector of $\Psi(1)$. This is equivalent to the requirement that

$$\Gamma_{\Psi(1)} \in \Lambda^0_\Delta$$

i.e., $\Gamma_{\Psi(1)}$ does not lie in the closure of the Maslov pseudo-cocycle. We are now in position to define the Conley–Zehnder index

Definition 10.4.1. Assume that $\Psi\colon [0,1] \to \mathrm{Sp}(V)$ is a non-degenerate path of symplectic linear maps starting at the identity. Then the *Conley–Zehnder index* of Ψ is defined as

$$\mu_{CZ}(\Psi) := \mu_\Delta(\Gamma_\Psi) \in \mathbb{Z}.$$

Remark 10.4.2. The Maslov index $\mu_\Delta(\Gamma_\Psi)$ is also defined in the case where Ψ is degenerate. However, in the degenerate case we do not define the Conley–Zehnder index via the Maslov index. Instead, we follow [126] and use the spectral flow to associate a Conley–Zehnder index to a degenerate path in Definition 11.2.2. We point out that in the case of degenerate paths the Conley–Zehnder index might in fact be different from the Maslov index. Indeed, the Conley–Zehnder index as extended to degenerate paths via Definition 11.2.2 becomes lower semi-continuous whereas the Maslov index is neither lower nor upper semi-continuous.

Since the Maslov index is in general only half integer valued it is a priori not clear that the Conley–Zehnder index takes values in the integers. However, this can be seen as follows. Since by assumption the path is non-degenerate, it follows that $\mathrm{sign} C(\Gamma_\Psi, \Delta, 1) = 0$. Therefore we obtain the formula

$$\mu_{CZ}(\Psi) = \frac{1}{2}\mathrm{sign} C(\Gamma_\Psi, \Delta, 0) + \sum_{0<t<1} \mathrm{sign} C(\Gamma_\Psi, \Delta, t). \tag{10.14}$$

Since $\Psi(0) = \mathrm{id}$ we have $\Gamma_\Psi(0) = \Delta$ and therefore $C(\Gamma_\Psi, \Delta, 0)$ is a quadratic form on the vector space Δ. However, the diagonal Δ is even-dimensional and therefore $\mathrm{sign} C(\Gamma_\Psi, \Delta, 0) \in 2\mathbb{Z}$. This proves that the Conley–Zehnder index is an integer.

10.5 Invariants of the group $SL(2,\mathbb{R})$

In the special case where the symplectic path takes values in the group $Sp(1)$ the Conley–Zehnder index can be computed with the help of the rotation number. In order to introduce the rotation number in the next section we need to have a closer look at the group $Sp(1)$. Because in two dimensions symplectic means just area preserving, the group $SL(2,\mathbb{R})$ corresponds to the group $Sp(1)$. The group $SL(2,\mathbb{R})$ acts on itself by conjugation. We are interested in the orbits of this action.

Namely if $A, B \in SL(2,\mathbb{R})$, we want to know if there exists $C \in SL(2,\mathbb{R})$ such that $A = CBC^{-1}$. This is more restrictive than the similarity relation studied in Linear Algebra where C is allowed to be chosen in $GL(2,\mathbb{R})$. In particular, quantities invariant under a similarity transformation are also invariants of an $SL(2,\mathbb{R})$ orbit. Such quantities are the determinant and the trace. Because all elements of $SL(2,\mathbb{R})$ have determinant one just the trace remains. To describe an additional invariant of an $SL(2,\mathbb{R})$ orbit which is not an invariant under similarity transformations we first introduce the *signum function*

$$\sigma\colon \mathbb{R} \to \{-1, 0, 1\}$$

defined for $r \in \mathbb{R}$ by

$$\sigma(r) = \begin{cases} 1 & r > 0 \\ 0 & r = 0 \\ -1 & r < 0. \end{cases}$$

We can now formulate the following lemma.

Lemma 10.5.1. *Assume that* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R})$ *satisfies* $|\mathrm{tr}(A)| \le 2$*. Then the quantity*

$$\varsigma(A) := \sigma(b - c)$$

is invariant under conjugation of A with matrices in $SL(2,\mathbb{R})$. Moreover, if the trace satisfies the strict inequality $|\mathrm{tr}(A)| < 2$, then $\varsigma(A) \in \{-1, 1\}$.

Example 10.5.2. For an angle $\theta \in (0,\pi)$ consider the rotations about θ and $-\theta$

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad R_{-\theta} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Then

$$\varsigma(R_\theta) = -1, \quad \varsigma(R_{-\theta}) = 1$$

so that we conclude that R_θ and $R_{-\theta}$ are not conjugated in $SL(2,\mathbb{R})$ although they are conjugated in $GL(2,\mathbb{R})$ by reflection at the x-axis.

Proof of Lemma 10.5.1*.* Suppose that $C = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \in SL(2,\mathbb{R})$. Abbreviate

$$\begin{aligned} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} := CAC^{-1} &= \begin{pmatrix} e & f \\ g & h \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} h & -f \\ -g & e \end{pmatrix} \\ &= \begin{pmatrix} aeh - beg + cfh - dfg & -aef + be^2 - cf^2 + def \\ agh - bg^2 + ch^2 - dgh & -afg + beg - cfh + deh \end{pmatrix}. \end{aligned}$$

We discuss two cases separately.

Case 1: $|a + d| < 2$.

In this case we estimate

$$0 \leq (a-d)^2 = a^2 - 2ad + d^2 = (a+d)^2 - 4ad < 4 - 4ad$$

which implies that

$$ad < 1.$$

Using the fact that the determinant of A is one we obtain from this the inequality

$$bc < 0. \tag{10.15}$$

This tells us that neither b nor c vanishes, and that they have different sign, so that

$$\varsigma(A) \in \{-1, 1\}.$$

We next show that

$$\sigma(b) = \sigma(b'), \quad \sigma(c) = \sigma(c'). \tag{10.16}$$

Because $\det(CAC^{-1}) = \det(A)$ and $\mathrm{tr}(CAC^{-1}) = \mathrm{tr}(A)$, the same reasoning which led to (10.15) gives us the inequality

$$b'c' < 0$$

so that again neither b' nor c' vanishes. Because the group $SL(2,\mathbb{R})$ is connected, we obtain (10.16). Because b and c have different sign, it holds that

$$\varsigma(A) = \sigma(b) = \sigma(b') = \varsigma(CAC^{-1}).$$

This finishes the proof of Case 1.

Case 2: $|a+d| = 2$.

In this case the strict inequality (10.15) has to be replaced by

$$bc \leq 0. \tag{10.17}$$

We distinguish the following subcases

Case 2a: $\varsigma(A) = 0$.

This means that

$$b = c$$

and therefore in view of (10.17) it holds that

$$b = c = 0.$$

Therefore A is a diagonal matrix. Moreover, in view of

$$ad = 1, \quad |a+d| = 2$$

we compute that

$$a = d = \pm 1$$

so that one of the following two cases holds

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In both cases we obtain

$$CAC^{-1} = A$$

so that trivially

$$\varsigma(A) = \varsigma(CAC^{-1}).$$

This proves Case 2a.

Case 2b: $\varsigma(A) \neq 0$.

We need to exclude a change of signs. Because $SL(2,\mathbb{R})$ is connected, it suffices to rule out the case that $b' - c' = 0$ or equivalently that $b' = c'$. This follows from the proof of Step 2a. Indeed, if $b' = c'$ then CAC^{-1} is just $\pm\mathrm{id}$ and then $A = \pm\mathrm{id}$ as well, contradicting the assumption that $\varsigma(A) \neq 0$. This finishes the proof of Case 2b and hence of the lemma. □

Theorem 10.5.3. *The conjugation class* $[A]$ *of a matrix* $A \in SL(2,\mathbb{R})$ *is uniquely determined by* $\mathrm{tr}(A)$ *if* $|\mathrm{tr}(A)| > 2$ *and by* $\mathrm{tr}(A)$ *together with* $\varsigma(A)$ *if* $|\mathrm{tr}(A)| \leq 2$.

Proof. We first consider the hyperbolic case, namely $|\mathrm{tr}(A)| > 2$. Because the determinant of A is one, we conclude that A has two real eigenvalues r and $\frac{1}{r}$ for $|r| > 1$. In particular, there exists $C \in GL(2,\mathbb{R})$ such that

$$CAC^{-1} = \begin{pmatrix} r & 0 \\ 0 & \frac{1}{r} \end{pmatrix}.$$

After rescaling C by a scalar we can assume without loss of generality that $\det(C) = \pm 1$. After possibly composing C with a reflection at the x-axis or y-axis, we can achieve that $\det(C) = 1$. This proves that the trace is a complete invariant of the conjugation class of A in case $|\mathrm{tr}(A)| > 2$.

We next consider the elliptic case, namely $|\mathrm{tr}(A)| < 2$. In this case A has two complex conjugated eigenvalues which both lie on the unit circle. Therefore there exists $C \in GL(2,\mathbb{R})$ such that

$$CAC^{-1} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where the angle θ satisfies

$$\cos(\theta) = \tfrac{\mathrm{tr}(A)}{2}.$$

Again by rescaling C with a scalar we can assume without loss of generality that $\det(C) = \pm 1$. In case the determinant of C equals -1 we compose C with a reflection at the x-axis and replace the angle θ by $-\theta$. Because by Lemma 10.5.1 we know that $\varsigma(A)$ is conjugation invariant in the case $|\mathrm{tr}(A)| \leq 2$, the theorem follows in the elliptic case.

It remains to discuss the parabolic case $|\mathrm{tr}(A)| = 2$. In this case A has just one eigenvalue which equals ± 1 of algebraic multiplicity two. If its geometric multiplicity equals two as well, then A is just $\pm\mathrm{id}$ and each matrix conjugated to A coincides with A. Therefore we are left with the case that the geometric multiplicity of the eigenvalue of A equals one. We restrict our discussion to the case where the eigenvalue of A is 1, the case -1 is completely analogous. In this case there exists $C \in GL(2,\mathbb{R})$ such that A is similar to a Jordan block

$$CAC^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Again we can assume that $\det(C) = \pm 1$. If $\det(C) = -1$ we replace C by its composition with a reflection at the x-axis or y-axis and obtain

$$CAC^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

This finishes the proof in the parabolic case and the theorem follows. □

Using the fact that the trace is a continuous function on $SL(2,\mathbb{R})$ and the only point at which the signum function is not continuous is zero, one sees that all orbits of the $SL(2,\mathbb{R})$-action on itself by conjugation are closed except the orbits

$$\left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right], \left[\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\right], \left[\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}\right], \quad \text{and} \quad \left[\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}\right]$$

which contain the Jordan block matrices. Indeed, the identity matrix lies in the closure of the first two orbits and minus the identity matrix lies in the closure of the second two orbits. To see that, note that for every $\epsilon > 0$ it holds that

$$\begin{pmatrix} \epsilon & 0 \\ 0 & \frac{1}{\epsilon} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\epsilon} & 0 \\ 0 & \epsilon \end{pmatrix} = \begin{pmatrix} 1 & \epsilon^2 \\ 0 & 1 \end{pmatrix}$$

and similarly for the other Jordan block matrices. In view of this description we see that the orbit closure relation $\sim$ on $SL(2,\mathbb{R})$ given by

$$A \sim B \quad \Longleftrightarrow \quad \overline{[A]} \cap \overline{[B]} \neq \emptyset$$

is actually an equivalence relation. We abbreviate by

$$\mathfrak{Q} := SL(2,\mathbb{R})/\sim$$

the set of equivalence classes. This is very reminiscent of the notion of a "good quotient" in Geometric Invariant Theory (GIT), see [191, 193], so that the reader might think of $\mathfrak{Q}$ as

$$\mathfrak{Q} = SL(2,\mathbb{R})//SL(2,\mathbb{R})$$

for the action of $SL(2,\mathbb{R})$ on itself by conjugation. Different from the "geometric quotient" $SL(2,\mathbb{R})/SL(2,\mathbb{R})$ the "good quotient" $SL(2,\mathbb{R})//SL(2,\mathbb{R})$ is Hausdorff. The difference is that the three distinct points

$$\left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right], \left[\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\right], \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \in SL(2,\mathbb{R})/SL(2,\mathbb{R})$$

as well as

$$\left[\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}\right], \left[\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}\right], \left[\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right] \in SL(2,\mathbb{R})/SL(2,\mathbb{R})$$

are identified to a single point in $\mathfrak{Q}$. In other words all matrices with trace 2 are identified in $\mathfrak{Q}$ while they give rise to three orbits in $SL(2,\mathbb{R})/SL(2,\mathbb{R})$ distinguished by the signum ς and analogously for matrices with trace -2. We write $[[A]]$ for an equivalence class of a matrix $A \in SL(2,\mathbb{R})$ in $\mathfrak{Q}$. We then have

$$[[A]] = [A], \quad |\mathrm{tr}(A)| \neq 2$$

and

$$\left[\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\right] = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] \cup \left[\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right] \cup \left[\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\right]$$

as well as

$$\left[\left[\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right]\right] = \left[\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right] \cup \left[\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}\right] \cup \left[\begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}\right].$$

Topologically the quotient $\mathfrak{Q}$ is a circle with two spikes, see Figure 10.2, and we can identify it with the set

$$\big\{z \in \mathbb{C} : ||z|| = 1\big\} \cup \big\{r \in \mathbb{R} : r \leq -1\big\} \cup \big\{r \in \mathbb{R} : r \geq 1\big\} \subset \mathbb{C}.$$

This identification can be achieved in the following way. For an element $e^{i\theta}$ on the unit circle in $\mathbb{C}$ we set

$$s(e^{i\theta}) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \in SL(2,\mathbb{R})$$

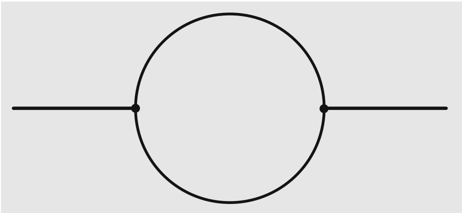


Figure 10.2: A spiked circle.

and for a real number r satisfying $|r| \geq 1$ we set

$$s(r) = \left(\begin{array}{cc} r & 0 \\ 0 & \frac{1}{r} \end{array} \right) \in SL(2, \mathbb{R}).$$

We then identify $\mathbb{Q}$ with the subset of $\mathbb{C}$ introduced above via the map

$$\big\{z \in \mathbb{C} : ||z|| = 1\big\} \cup \big\{r \in \mathbb{R} : r \leq -1\big\} \cup \big\{r \in \mathbb{R} : r \geq 1\big\} \to \mathfrak{Q}, \quad z \mapsto [[s(z)]].$$

Using this identification, we can think of the map s actually as a section from the quotient $\mathfrak{Q} = SL(2, \mathbb{R})//SL(2, \mathbb{R})$ back to the space $SL(2, \mathbb{R})$. In particular, we have verified the nontrivial fact that such a section exists. In the following, we use this section to also identify $\mathfrak{Q}$ with a subset of $SL(2, \mathbb{R})$. Geometrically, the unit circle in $\mathbb{C}$ minus the points ± 1 corresponds to equivalence classes of elliptic matrices in $SL(2, \mathbb{R})$, the spikes minus the points ± 1 to equivalence classes of hyperbolic matrices, whereas the points ± 1 correspond to equivalence classes of parabolic matrices.

We further decompose

$$\mathfrak{Q} = \mathfrak{Q}_e \cup \mathfrak{Q}_h^- \cup \mathfrak{Q}_h^+ \tag{10.18}$$

where

$$\mathfrak{Q}_e = \{z \in \mathbb{C} : ||z|| = 1\}$$

is the unit circle in $\mathbb{C}$ which corresponds to the elliptic matrices (including the parabolic ones),

$$\mathfrak{Q}_h^+ = \{r \in \mathbb{R} : r \geq 1\}$$

which is the positive hyperbolic spike and

$$\mathfrak{Q}_h^- = \{r \in \mathbb{R} : r \leq -1\}$$

which is the negative hyperbolic spike. If we think of $\mathfrak{Q}$ as a subset of $SL(2, \mathbb{R})$ using the section s described above, then

$$\mathfrak{Q}_e = SO(2) \subset SL(2, \mathbb{R})$$

are the rotations.

10.6 The rotation number

Let $\Psi \colon [0, 1] \to SL(2, \mathbb{R}) = Sp(1)$ be a smooth path of symplectic matrices with the property that $\Psi(0) = \mathrm{id}$. Let

$$\pi_1 \colon SL(2, \mathbb{R}) \to \mathfrak{Q} = SL(2, \mathbb{R})//SL(2, \mathbb{R}), \quad A \mapsto [[A]]$$

be the canonical projection. Using the decomposition (10.18) we further obtain a retraction

$$\pi_2 \colon \mathfrak{Q} \to \mathfrak{Q}_e$$

which is defined by

$$\pi_2|_{\mathfrak{Q}_e} = \mathrm{id}|_{\mathfrak{Q}_e}, \quad \pi_2(r) = 1,\ r \in \mathfrak{Q}_h^+, \quad \pi_2(r) = -1,\ r \in \mathfrak{Q}_h^-,$$

i.e., π_2 collapses the two spikes. Note that $\mathfrak{Q}_e$ is just the unit circle in $\mathbb{C}$ which we identify with $S^1 = \mathbb{R}/\mathbb{Z}$ via the map $r \mapsto e^{2\pi i r}$. Composing these two maps we obtain a map

$$\pi := \pi_2 \circ \pi_1 \colon SL(2, \mathbb{R}) \to S^1.$$

In particular, if we compose Ψ with π we get a map

$$\pi \circ \Psi \colon [0, 1] \to S^1.$$

We lift this map to the universal cover $\mathbb{R}$ of $S^1 = \mathbb{R}/\mathbb{Z}$ to get a map

$$\widetilde{\pi \circ \Psi} \colon [0, 1] \to \mathbb{R}$$

with the property that

$$\widetilde{\pi \circ \Psi}(0) = 0.$$

This is possible, since Ψ starts at the identity by assumption. We now define the rotation number of the path of symplectic matrices Ψ as

$$\mathrm{rot}(\Psi) := \widetilde{\pi \circ \Psi}(1) \in \mathbb{R}.$$

Note that if Ψ_0 and Ψ_1 are two paths of symplectic matrices starting at the identity which are homotopic with fixed endpoints or more generally homotopic such that during the homotopy the endpoint stays in a fixed fiber of the projection $\pi_1 \colon SL(2, \mathbb{R}) \to \mathfrak{Q}$, then

$$\mathrm{rot}(\Psi_0) = \mathrm{rot}(\Psi_1).$$

Recall that the path Ψ is called non-degenerate if $\det(\Psi(1) - \mathrm{id}) \neq 0$, i.e., $\Psi(1)$ has no eigenvalue equal to one. Because in this section Ψ takes values in $SL(2, \mathbb{R}) = Sp(1)$, this can equivalently be rephrased as $\mathrm{tr}(\Psi(1)) \neq 2$. We abbreviate by

$$\lfloor \cdot \rfloor,\ \lceil \cdot \rceil \colon \mathbb{R} \to \mathbb{Z}$$

the floor and ceiling function defined for $r \in \mathbb{R}$ as

$$\lfloor r \rfloor = \max\{n \in \mathbb{Z} : n \leq r\}, \quad \lceil r \rceil = \min\{n \in \mathbb{Z} : n \geq r\}.$$

We can now explain how the rotation number determines the Conley–Zehnder index of the path Ψ.

Theorem 10.6.1. *Assume that $\Psi\colon [0,1] \to SL(2,\mathbb{R})$ is a non-degenerate path of symplectic matrices starting at the identity. Then the Conley–Zehnder index of Ψ can be computed as follows.*

(i) *Assume that $|\mathrm{tr}(\Psi(1))| \leq 2$, i.e, $\Psi(1)$ is elliptic or parabolic. then*

$$\mu_{CZ}(\Psi) = 2\lfloor \mathrm{rot}(\Psi) \rfloor + 1 = 2\lceil \mathrm{rot}(\Psi) \rceil - 1 = \lfloor \mathrm{rot}(\Psi) \rfloor + \lceil \mathrm{rot}(\Psi) \rceil.$$

(ii) *Assume that $|\mathrm{tr}(\Psi(1))| > 2$, so in other words $\Psi(1)$ is hyperbolic and therefore $\mathrm{rot}(\Psi) \in \frac{1}{2}\mathbb{Z}$, then*

$$\mu_{CZ}(\Psi) = 2\mathrm{rot}(\Psi).$$

Before embarking on its proof, we draw two corollaries from this theorem.

Corollary 10.6.2. *Assume that Ψ is as in Theorem 10.6.1. If $\Psi(1)$ is elliptic or parabolic, then $\mu_{CZ}(\Psi)$ is odd. If $\Psi(1)$ is positively hyperbolic, i.e., $\mathrm{tr}(\Psi(1)) > 2$ or equivalently $\Psi(1)$ has two positive real eigenvalues, then $\mu_{CZ}(\Psi)$ is even. Finally, if $\Psi(1)$ is negatively hyperbolic, i.e., $\mathrm{tr}(\Psi(1)) < -2$ or equivalently $\Psi(1)$ has two negative real eigenvalues, then $\mu_{CZ}(\Psi)$ is odd again.*

Corollary 10.6.3. *Assume that Ψ is as in Theorem 10.6.1. If $\mu_{CZ}(\Psi) \geq 3$, then $\mathrm{rot}(\Psi) > 1$.*

Proof of Theorem 10.6.1. In the proof we identify $\mathfrak{Q}$ with a subset of $SL(2,\mathbb{R})$ via the map s. We first consider the case where $\Psi(t) \in \mathfrak{Q}$ for every $t \in [0,1]$. Note that $\mathfrak{Q}$ intersects the Maslov cycle precisely in the point 1. Therefore it suffices to understand the contribution in the crossing formula if the path Ψ crosses 1. To describe this we distinguish between the upper and lower elliptic branch

$$\mathfrak{Q}_e^+ := \big\{z = e^{2\pi i\theta} \in \mathfrak{Q}_e : \theta \in (0,\pi)\big\}, \quad \mathfrak{Q}_e^- := \big\{z = e^{2\pi i\theta} \in \mathfrak{Q}_e : \theta \in (-\pi,0)\big\}$$

Note that the upper elliptic branch $\mathfrak{Q}_e^+$ corresponds to orthogonal matrices $A \in SO(2)$ satisfying $\varsigma(A) = 1$ and $|\mathrm{tr}(A)| \neq 2$ and the negative elliptic branch $\mathfrak{Q}_e^-$ corresponds to orthogonal matrices $A \in SO(2)$ meeting $\varsigma(A) = -1$ and $|\mathrm{tr}(A)| \neq 2$. If the crossing goes from $\mathfrak{Q}_e^-$ to the positive hyperbolic branch $\mathfrak{Q}_h^+$, then the contribution to the crossing number is $+1$. If the crossing goes from $\mathfrak{Q}_h^+$ to $\mathfrak{Q}_e^+$, the crossing number is $+1$ again. Moreover, if the crossing goes directly from $\mathfrak{Q}_e^-$ to $\mathfrak{Q}_e^+$ the crossing number is $+2$. Of course the crossing number has to be multiplied by a minus if the crossing goes backwards. Using these rules the assertion of the theorem now follows immediately from the crossing formula for the Conley–Zehnder index (10.14) provided the path Ψ takes values in $\mathfrak{Q}$.

For the general case where the path Ψ is not supposed to take values in $\mathfrak{Q}$, note that in view of Lemma 10.6.4 below such a path is homotopic to a path in $\mathfrak{Q}$ for a homotopy with endpoint in a fixed fiber of the projection $\pi_1\colon SL(2,\mathbb{R}) \to \mathfrak{Q}$. Because the Conley–Zehnder index as well as the rotation number are invariant under such a homotopy the general case can be reduced to the special case just proved. $\square$

Lemma 10.6.4. *$SO(2)$ is a strong deformation retract of $SL(2,\mathbb{R})$.*

Proof. A retraction of $SL(2,\mathbb{R})$ to $SO(2)$ can be constructed using the polar decomposition. Namely the map

$$SL(2,\mathbb{R}) \times [0,1] \to SL(2,\mathbb{R}), \quad (A,r) \mapsto (AA^T)^{-r/2}A$$

does the job. $\square$

Remark 10.6.5. In fact, we know from hyperbolic geometry that the unit tangent bundle of the hyperbolic plane $\mathbb{H}^2$, which is clearly a trivial circle bundle, can be identified with $\mathbb{P}SL(2,\mathbb{R})$, so $SL(2,\mathbb{R})$ is diffeomorphic (but not isomorphic as a Lie group) to $S^1 \times \mathbb{R}^2$.

Chapter 11

Spectral Flow

11.1 A Fredholm operator and its spectrum

In the following we let ω denote the standard symplectic form on $\mathbb{C}^n$. We abbreviate by $\mathrm{Sp}(n)$ the linear symplectic group consisting of all real linear transformations $A\colon \mathbb{C}^n \to \mathbb{C}^n$ satisfying $A^*\omega = \omega$.

Suppose that $\Psi\colon [0,1] \to \mathrm{Sp}(n)$ is a smooth path of symplectic matrices which starts at the identity, i.e., $\Psi(0) = \mathrm{id}$. It follows that $\Psi'(t)\Psi^{-1}(t) \in \mathrm{Lie}\,\mathrm{Sp}(n)$, the Lie algebra of the linear symplectic group. The Lie algebra of the linear symplectic group can be described as follows. For the splitting $\mathbb{C}^n = \mathbb{R}^n \times \mathbb{R}^n$ write

$$J = \begin{pmatrix} 0 & -\mathrm{id} \\ \mathrm{id} & 0 \end{pmatrix}. \tag{11.1}$$

Note that

$$J^2 = -\mathrm{id}, \quad J^T = -J.$$

Using J, the condition that a matrix A lies in $Sp(n)$ is equivalent to the assertion that A satisfies

$$A^T J A = J.$$

That means that $B \in \mathrm{Lie}\,\mathrm{Sp}(n)$ if and only if

$$B^T J + JB = 0.$$

Therefore

$$(JB)^T = B^T J^T = -B^T J = JB$$

implying that $JB \in \mathrm{Sym}(2n)$, the vector space of symmetric $2n \times 2n$-matrices. On the other hand one checks immediately that if $JB \in \mathrm{Sym}(2n)$, then $B \in \mathrm{Lie}\,Sp(n)$, which means that the map $B \mapsto JB$ is a vector space isomorphism between $\mathrm{Lie}\,\mathrm{Sp}(n)$ and $\mathrm{Sym}(2n)$.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_11

Therefore to any smooth path $\Psi \in C^\infty([0,1], \mathrm{Sp}(n))$ satisfying $\Psi(0) = \mathrm{id}$ we associate a smooth path $S = S_\Psi \in C^\infty([0,1], \mathrm{Sym}(2n))$ by setting

$$S(t) := -J\Psi'(t)\Psi^{-1}(t). \tag{11.2}$$

We recover Ψ from S by solving the ODE

$$\Psi'(t) = JS(t)\Psi(t),\ t \in [0,1], \quad \Psi(0) = \mathrm{id}. \tag{11.3}$$

That means we have a one-to-one correspondence between paths of linear symplectic matrices starting at the identity and paths of symmetric matrices. We associate to these a linear operator

$$A = A_\Psi = A_S \colon W^{1,2}(S^1, \mathbb{C}^n) \to L^2(S^1, \mathbb{C}^n), \quad v \mapsto -J\partial_t v - Sv. \tag{11.4}$$

Here L^2 refers to the Hilbert space of square integrable functions and $W^{1,2}$ refers to the Hilbert space of square integrable functions which admit a weak derivative which is also square integrable. As a reminder to the reader, if $u \in L^1(S^1, \mathbb{R})$ is an absolutely integrable function, then we call $v \in L^1(S^1, \mathbb{R})$ a *weak derivative* of u if

$$\int_{S^1} u(t)\phi'(t)dt = -\int_{S^1} v(t)\phi(t)dt$$

for every $\phi \in C^\infty(S^1, \mathbb{R})$.

Our goal is to relate the spectral theory of the operators A_Ψ to the Conley–Zehnder index of Ψ. For that purpose we first examine the kernel of A.

Lemma 11.1.1. *The evaluation map*

$$E\colon \ \ker A \to \ker(\Psi(1) - \mathrm{id}), \quad v \mapsto v(0)$$

is a vector space isomorphism.

Proof. We prove the lemma in three steps.

Step 1: The image of E lies in $\ker(\Psi(1) - \mathrm{id})$.

Pick $v \in \ker A$. We have to show that $v(0) \in \ker(\Psi(1) - \mathrm{id})$. The condition that $Av = 0$ means that v is a solution of the ODE

$$J\partial_t v = -Sv \tag{11.5}$$

or equivalently

$$\partial_t v = (\partial_t \Psi)\Psi^{-1} v = -\Psi\partial_t(\Psi^{-1})v.$$

This can be rephrased by saying that $\partial_t(\Psi^{-1}v) = 0$ or equivalently that

$$v(t) = \Psi(t)v_0, \quad v_0 \in \mathbb{C}^n.$$

Since v is a loop, we obtain

$$v(0) = v(1) = \Psi(1)v(0).$$

This proves that the evaluation map is well defined.

Step 2: The evaluation map is injective.

Suppose that $v \in \ker E$. That means that $v(0) = 0$. However, v is a solution of the ODE (11.5). This implies that $v(t) = 0$ for every $t \in S^1$.

Step 3: The evaluation map is surjective.

Suppose that $v_0 \in \ker(\Psi(1) - \mathrm{id})$, i.e.,

$$\Psi(1)v_0 = v_0.$$

Define

$$v(t) = \Psi(t)v_0.$$

It follows that

$$v(1) = \Psi(1)v_0 = v_0 = v(0).$$

Hence $v \in W^{1,2}(S^1, \mathbb{C}^n)$ and we have seen in the proof of Step 1 that $v \in \ker A$. This finishes the proof of Step 3 and hence of the lemma. □

It is worth pointing out that when $\Psi(1)$ has an eigenvector to the eigenvalue 1, this precisely means that its graph $\Gamma_{\Psi(1)}$ lies in the closure of the Maslov pseudocycle Λ^1_Δ. Hence by the lemma crossing the Maslov pseudocycle is equivalent to an eigenvalue of the operator crossing zero.

The L^2-inner product for two vectors $v_1, v_2 \in L^2(S^1, \mathbb{C}^n)$ is given by

$$\langle v_1, v_2 \rangle = \int_0^1 \langle v_1(t), v_2(t) \rangle dt = \int_0^1 \omega(v_1(t), Jv_2(t))dt.$$

Lemma 11.1.2. *The operator A is symmetric with respect to the L^2-inner product.*

Proof. Suppose that $v_1, v_2 \in W^{1,2}(S^1, \mathbb{C}^n)$. We compute using integration by parts and the fact that S is symmetric

$$\begin{aligned}
\langle Av_1, v_2 \rangle &= -\int_0^1 \langle J\partial_t v_1 + Sv_1, v_2 \rangle dt \\
&= -\int_0^1 \omega(J\partial_t v_1, J\partial_t v_2\rangle dt - \int_0^1 \langle Sv_1, v_2 \rangle dt \\
&= -\int_0^1 \omega(\partial_t v_1, v_2)dt - \int_0^1 \langle v_1, Sv_2 \rangle dt \\
&= -\int_0^1 \omega(v_1, \partial_t v_2)dt - \int_0^1 \langle v_1, Sv_2 \rangle dt \\
&= \langle v_1, Av_2 \rangle.
\end{aligned}$$

This finishes the proof of the lemma. □

A crucial property of the operator $A\colon W^{1,2}(S^1,\mathbb{C}^n) \to L^2(S^1,\mathbb{C}^n)$ is that it is a Fredholm operator of index 0. Before stating this theorem we recall some basic facts about Fredholm operators without proofs. Proofs can be found for example in [173, Appendix A.1]. If H_1 and H_2 are Hilbert spaces, then a bounded linear operator $D\colon H_1 \to H_2$ is called a *Fredholm operator* if it has a finite-dimensional kernel, a closed image and a finite-dimensional cokernel. Then the number

$$\mathrm{ind}D := \dim\ker D - \dim \mathrm{coker}D \in \mathbb{Z}$$

is referred to as the *index* of the Fredholm operator D. For example if H_1 and H_2 are finite-dimensional, then every linear operator $D\colon H_1 \to H_2$ is Fredholm and its index is given by

$$\mathrm{ind}D = \dim H_1 - \dim H_2.$$

Interestingly in this example the Fredholm index does not depend on D at all, although $\dim\ker D$ and $\dim \mathrm{coker}D$ do. The usefulness of Fredholm operators lies in the fact that similar phenomena happen in infinite dimensions and the Fredholm index is rather stable under perturbations.

Recall that a *compact operator* $K\colon H_1 \to H_2$ is a bounded linear operator with the property that the closure $\overline{K(B_{H_1})} \subset H_2$ is compact, where $B_{H_1} = \{v \in H_1 : ||v|| \leq 1\}$ is the unit ball in H_1 and the closure refers to the topology in H_2. The first useful fact about Fredholm operators is that they are stable under compact perturbations.

Theorem 11.1.3. *Assume that $D\colon H_1 \to H_2$ is a Fredholm operator and $K\colon H_1 \to H_2$ is a compact operator. Then $D+K\colon H_1 \to H_2$ is a Fredholm operator as well and*

$$\mathrm{ind}(D+K) = \mathrm{ind}D.$$

The second useful fact about Fredholm operators is that they are stable under small perturbations in the operator topology.

Theorem 11.1.4. *Assume that $D\colon H_1 \to H_2$ is a Fredholm operator. Then there exists $\epsilon > 0$ such that for every $E\colon H_1 \to H_2$ satisfying $||D - E|| < \epsilon$, where the norm refers to the operator norm, it holds that E is still Fredholm and* $\mathrm{ind}E = \mathrm{ind}D$.

In order to prove that an operator is Fredholm the following lemma is very useful.

Lemma 11.1.5. *Assume that H_1, H_2, H_3 are Hilbert spaces, $D\colon H_1 \to H_2$ is a bounded linear operator, $K\colon H_1 \to H_3$ is a compact operator, and there exists a constant $c > 0$ such that the following estimate holds for every $x \in H_1$*

$$||x||_{H_1} \leq c\big(||Dx||_{H_2} + ||Kx||_{H_3}\big).$$

Then D has a closed image and a finite-dimensional kernel.

A bounded linear operator with a closed image and a finite-dimensional kernel is referred to as a *semi Fredholm operator*. Therefore under the conditions of the lemma, the operator D is a semi Fredholm operator. In practice, one can often apply the lemma again to the adjoint of D. Since the kernel of the adjoint of D coincides with the cokernel of D, this enables one to deduce the Fredholm property of D. As we mentioned already, the proofs of the two previous theorems as well as of the lemma can be found in [173, Appendix A.1.]. We can use these results to prove the following theorem.

Theorem 11.1.6. *The operator* $A\colon W^{1,2}(S^1,\mathbb{C}^n) \to L^2(S^1,\mathbb{C}^n)$ *is a Fredholm operator of index* 0.

We present two proofs of this theorem.

Proof 1 *of Theorem* 11.1.6. Taking advantage of the fact that the inclusion of $W^{1,2}(S^1,\mathbb{C}^n) \hookrightarrow L^2(S^1,\mathbb{C}^n)$ is compact we note that the operator $A_S = -J\partial_t + S$ is a compact perturbation of the operator $A_0 = -J\partial_t$. To check that the operator A_0 is Fredholm of index 0 is a straightforward exercise in the fundamental theorem of calculus and is left to the reader. Now it follows from Theorem 11.1.3 that A_S is Fredholm of index 0 as well. □

In the second proof we use Lemma 11.1.5 instead of Theorem 11.1.3. The reason why we present it is that according to a similar scheme many elliptic operators can be proven to be Fredholm operators. These proofs usually contain two ingredients. First, one needs to produce an estimate in order to be able to apply Lemma 11.1.5. Then one has to apply elliptic regularity in order to identify the cokernel of the operator with the kernel of the adjoint. Since the operator A is an operator in just one variable the elliptic regularity part is immediate. The estimate is the content of the following lemma.

Lemma 11.1.7. *There exists* $c > 0$ *such that for every* $v \in W^{1,2}(S^1,\mathbb{C}^n)$ *the following estimate holds*

$$||v||_{W^{1,2}} \leq c\big(||v||_{L^2} + ||Av||_{L^2}\big).$$

Proof. Since $J\partial_t v = -Av - Sv$ we obtain

$$\partial_t v = JAv + JSv.$$

Therefore we estimate

$$\begin{aligned}
||v||^2_{W^{1,2}} &= ||v||^2_{L^2} + ||\partial_t v||^2_{L^2} \\
&= ||v||^2_{L^2} + ||J(A+S)v||^2_{L^2} \\
&= ||v||^2_{L^2} + ||(A+S)v||^2_{L^2} \\
&= ||v||^2_{L^2} + ||Av||^2_{L^2} + 2\langle Av, Sv\rangle + ||Sv||^2_{L^2} \\
&\leq ||v||^2_{L^2} + 2||Av||^2_{L^2} + 2||Sv||^2_{L^2} \\
&\leq c\big(||v||^2_{L^2} + ||Av||^2_{L^2}\big).
\end{aligned}$$

This finishes the proof of the lemma. □

Proof 2 of Theorem 11.1.6. The inclusion $W^{1,2}(S^1,\mathbb{C}^n) \hookrightarrow L^2(S^1,\mathbb{C}^n)$ is compact. Therefore, we deduce from Lemma 11.1.5 and Lemma 11.1.7 that A is a semi Fredholm operator, i.e., $\ker A$ is finite-dimensional and $\operatorname{im} A$ is closed. To determine its cokernel, choose $w \in \operatorname{im} A^\perp$, the orthogonal complement of the image of A. That means that $w \in L^2(S^1,\mathbb{C}^n)$ satisfies

$$\langle w, Av\rangle = 0, \quad \forall\ v \in W^{1,2}(S^1,\mathbb{C}^n).$$

Hence if $v \in W^{1,2}(S^1,\mathbb{C}^n)$, we have

$$0 = \langle w, Sv\rangle + \langle w, J\partial_t v\rangle = \langle Sw, v\rangle - \langle Jw, \partial_t v\rangle$$

and therefore

$$\langle Jw, \partial_t v\rangle = \langle Sw, v\rangle, \quad \forall\ v \in W^{1,2}(S^1,\mathbb{C}^n).$$

This implies that w, which a priori was just an element in $L^2(S^1,\mathbb{C}^n)$, actually lies in $W^{1,2}(S^1,\mathbb{C}^n)$ and satisfies the equation

$$J\partial_t w = -Sw.$$

In particular,

$$Aw = 0.$$

Since the operator A is symmetric by Lemma 11.1.2, it holds that every element in the kernel of A is orthogonal to the image of A. We obtain

$$\operatorname{im} A^\perp = \ker A$$

and therefore

$$\dim \operatorname{coker} A = \dim \ker A.$$

In particular, A is Fredholm and its index satisfies

$$\operatorname{ind} A = \dim \ker A - \dim \operatorname{coker} A = 0.$$

This proves Theorem 11.1.6 again. □

Remark 11.1.8. It follows in particular from Theorem 11.1.6 that the dimension of the kernel of the operator A is finite. In fact, actually more can be said, namely

$$\dim \ker A \leq 2n.$$

Indeed, since elements of the kernel satisfy a linear ODE, linearly independent elements of the kernel must be pointwise linearly independent. Therefore, if the kernel had dimension larger than $2n$, one could choose $2n+1$ linearly independent elements and use them to construct a nonzero element of the kernel that vanished at a point, since these elements are maps to $\mathbb{C}^n$. But since an element of the kernel that vanishes at a point vanishes identically, this is a contradiction.

The fact that $A\colon W^{1,2}(S^1,\mathbb{C}^n) \to L^2(S^1,\mathbb{C}^n)$ is a symmetric Fredholm operator of index 0 has important consequences for its spectrum. A reader familiar with the theory of unbounded operators might recognize that the fact that A is Fredholm of index 0 implies that A interpreted as an unbounded operator $A\colon L^2(S^1,\mathbb{C}^n) \to L^2(S^1,\mathbb{C}^n)$ is *self-adjoint* with dense domain $W^{1,2}(S^1,\mathbb{C}^n) \subset L^2(S^1,\mathbb{C}^n)$. Here we do not invoke the theory of self-adjoint unbounded operators but argue directly via the standard properties of Fredholm operators we already recalled. We abbreviate by

$$I\colon W^{1,2}(S^1,\mathbb{C}^n) \to L^2(S^1,\mathbb{C}^n)$$

the inclusion, which is a compact operator, and define the *spectrum* of A as

$$\mathfrak{S}(A) := \big\{\eta \in \mathbb{C} : A - \eta I \text{ not invertible}\big\}.$$

Its complement is the *resolvent set*

$$\mathfrak{R}(A) := \mathbb{C} \setminus \mathfrak{S}(A) = \big\{\eta \in \mathbb{C} : A - \eta I \text{ invertible}\big\}.$$

An element $\eta \in \mathbb{C}$ is called an *eigenvalue* of A if $\ker(A - \eta I) \neq \{0\}$.

Lemma 11.1.9. *The spectrum of A consists precisely of the eigenvalues of A, i.e.,*

$$\eta \in \mathfrak{S}(A) \iff \eta \text{ eigenvalue of } A.$$

Proof. The implication "$\Longleftarrow$" is obvious. To prove the implication "$\Longrightarrow$", we observe that, in view of the stability of the Fredholm index under compact perturbations as stated in Theorem 11.1.3, it follows that $A-\eta I$ is still a Fredholm operator of index 0. Now assume that $\eta \in \mathbb{C}$ is not an eigenvalue of A, i.e., $\ker(A-\eta I) = \{0\}$. Since the index of $A - \eta I$ is 0, we conclude that $\operatorname{coker}(A - \eta I) = \{0\}$. This means that $A - \eta I$ is bijective and hence by the open mapping theorem invertible. This finishes the proof of the lemma. $\square$

Lemma 11.1.10. *The spectrum is real, i.e., $\mathfrak{S}(A) \subset \mathbb{R}$.*

Proof. Assume that $\eta \in \mathfrak{S}(A)$. By Lemma 11.1.9 we know that η is an eigenvalue of A. Now the argument is standard. Indeed, if $v \neq 0$ such that $Av = \eta I v$, then we have the following in view of the symmetry of A established in Lemma 11.1.2,

$$\eta||v||^2 = \langle Av, v\rangle = \langle v, Av\rangle = \bar{\eta}||v||^2,$$

and therefore since $||v||^2 \neq 0$, we conclude $\eta = \bar{\eta}$, or in other words $\eta \in \mathbb{R}$. $\square$

Lemma 11.1.11. *The spectrum $\mathfrak{S}(A)$ is countable.*

Proof. If $\eta \in \mathfrak{S}(A)$ it follows from Lemma 11.1.9 that there exists $e_\eta \neq 0$ such that $Ae_\eta = \eta I e_\eta$. Since A is symmetric, we have for $\eta \neq \eta' \in \mathfrak{S}(A)$ that $\langle e_\eta, e_{\eta'}\rangle = 0$, i.e., the two eigenvectors are orthogonal to each other. Since the Hilbert space $L^2(S^1,\mathbb{C}^n)$ is separable, we conclude that $\mathfrak{S}(A)$ has to be countable. $\square$

As a consequence of Lemma 11.1.11 we conclude that there exists $\zeta_0 \in \mathbb{R}$ such that $\zeta_0 \notin \mathfrak{S}(A)$. In particular, $A - \zeta_0 I$ is invertible. Its inverse is a bounded linear map

$$(A - \zeta_0 I)^{-1} \colon L^2(S^1, \mathbb{C}^n) \to W^{1,2}(S^1, \mathbb{C}^n).$$

We define the resolvent operator

$$R(\zeta_0) := I \circ (A - \zeta_0 I)^{-1} \colon L^2(S^1, \mathbb{C}^n) \to L^2(S^1, \mathbb{C}^n).$$

Since the inclusion operator $I \colon W^{1,2}(S^1, \mathbb{C}^n) \to L^2(S^1, \mathbb{C}^n)$ is compact, we conclude that the resolvent operator $R(\zeta_0)$ is compact. Moreover, since A is symmetric, the resolvent operator $R(\zeta_0)$ is symmetric as well. Furthermore, 0 is not an eigenvalue of $R(\zeta_0)$, because if $R(\zeta_0)v = 0$, we get $v = (A - \zeta_0 I)R(\zeta_0)v = 0$.

Recall that if H is a Hilbert space and $R \colon H \to H$ is a compact symmetric operator then the *spectral theorem for compact symmetric operators* tells us that the spectrum $\mathfrak{S}(R)$ is real, bounded and zero is the only possible accumulation point of $\mathfrak{S}(R)$. Moreover, for all $\eta \in \mathfrak{S}(R)$ there exist pairwise commuting orthogonal projections

$$\Pi_\eta \colon H \to H, \quad \Pi_\eta^2 = \Pi_\eta = \Pi_\eta^*,\ \Pi_\eta \Pi_{\eta'} = \Pi_{\eta'} \Pi_\eta$$

satisfying

$$\sum_{\eta \in \mathfrak{S}(R)} \Pi_\eta = \mathrm{id}$$

such that

$$R = \sum_{\eta \in \mathfrak{S}(R)} \eta \Pi_\eta.$$

Moreover, if $\eta \neq 0 \in \mathfrak{S}(R)$ then $\dim \operatorname{im} \Pi_\eta < \infty$, meaning that the eigenvalue η has finite multiplicity.

From the spectral decomposition of the resolvent $R(\zeta_0)$ we obtain the spectral decomposition of the operator A, namely

$$A = \sum_{\eta \in \mathfrak{S}(R(\zeta_0))} \left(\frac{1}{\eta} + \zeta_0 \right) \Pi_\eta.$$

In particular, we can improve Lemma 11.1.11 to the following stronger statement.

Proposition 11.1.12. *The spectrum $\mathfrak{S}(A) \subset \mathbb{R}$ is discrete.*

If we write for the spectral decomposition of A

$$A = \sum_{\eta \in \mathfrak{S}(A)} \eta \Pi_\eta$$

and $\zeta \in \mathfrak{R}(A)$ is in the resolvent set of A, then we obtain the spectral decomposition of the resolvent operator $R(\zeta)$ as

$$R(\zeta) = \sum_{\eta \in \mathfrak{S}(A)} \frac{1}{\eta - \zeta} \Pi_\eta.$$

Now choose a smooth loop $\Gamma\colon S^1 \to \mathfrak{R}(A) \subset \mathbb{C}$ such that the only eigenvalue of A encircled by Γ is η (with winding number 1). Now we can recover the projection Π_η by the residue theorem as follows

$$\Pi_\eta = \frac{i}{2\pi} \int_\Gamma R(\zeta) d\zeta.$$

This formula plays a central role in Kato's fundamental book [146]. According to [146] this formula was first used in perturbation theory by Szökevalvi-Nagy [232] and Kato [144, 145]. More generally, if $\Gamma\colon S^1 \to \mathfrak{R}(A)$ is a smooth loop and $\mathfrak{S}_\Gamma(A) \subset \mathfrak{S}(A)$ denotes the set of all eigenvalues of A encircled by Γ (again with winding number 1), we obtain by the residue theorem the following formula

$$\frac{i}{2\pi} \int_\Gamma R(\zeta) d\zeta = \sum_{\eta \in \mathfrak{S}_\Gamma(A)} \Pi_\eta.$$

The reason why this formula is so useful is that the map

$$A \mapsto \frac{i}{2\pi} \int_\Gamma R_A(\zeta) d\zeta$$

is continuous in the operator A with respect the operator topology. In particular, the eigenvalues of A vary continuously under perturbations of A.

11.2 The spectrum bundle

Recall that the operator $A = A_S\colon W^{1,2}(S^1, \mathbb{C}^n) \to L^2(S^1, \mathbb{C}^n)$ is given as $v \mapsto -J\partial_t v - S(t)v$. In particular, for $S = 0$ we have the map $A_0 = -J\partial_t$. Its eigenvalues are $2\pi\ell$ for $\ell \in \mathbb{Z}$ and the corresponding eigenvectors are given by $v_0 e^{2\pi i \ell t}$ for $v_0 \in \mathbb{C}^n$. In particular, each eigenvalue has multiplicity $2n$. Abbreviate by

$$\mathcal{P} := C^\infty([0,1], \mathrm{Sym}(2n))$$

the space of paths of symmetric $2n \times 2n$-matrices. We endow the space $\mathcal{P}$ with the metric

$$d(S, S') = \int_0^1 ||S(t) - S'(t)|| dt.$$

The *Spectrum bundle* is defined as

$$\mathfrak{S} := \{(S, \lambda) \in \mathcal{P} \times \mathbb{R} : \lambda \in \mathfrak{S}(A_S)\} \subset \mathcal{P} \times \mathbb{R}.$$

It comes with a canonical projection $\mathfrak{S} \to \mathcal{P}$. In view of the continuity of eigenvalues of the operators A_S under perturbation and the description of the spectrum of A_0 discussed above, there exist continuous sections

$$\eta_k \colon \mathcal{P} \to \mathfrak{S}, \quad k \in \mathbb{Z}$$

which are uniquely determined by the following requirements

(i) The map $S \mapsto \eta_k(S)$ is continuous for every $k \in \mathbb{Z}$.

(ii) $\eta_k(S) \leq \eta_{k+1}(S)$ for every $k \in \mathbb{Z}$, $S \in \mathcal{P}$.

(iii) $\mathfrak{S}(A_S) = \{\eta_k(S) : k \in \mathbb{Z}\}$.

(iv) If $\eta \in \mathfrak{S}(A_S)$, then the number $\#\{k \in \mathbb{Z} : \eta_k(S) = \eta\}$ equals the multiplicity of the eigenvalue η.

(v) The sections η_k are normalized such that $\eta_j(0) = 0$ for $j \in \{1, \dots, 2n\}$.

Recall that if $S \in \mathcal{P}$ we can associate to S a path of symplectic matrices Ψ_S starting at the identity by formula (11.3). The following theorem relates the Conley–Zehnder index of Ψ_S as defined in Definition 10.4.1 to the spectrum of A_S, see also [126, 212].

Theorem 11.2.1 (Spectral flow)**.** *Assume that* Ψ_S *is non-degenerate, meaning that* $\det(\Psi_S(1) - \mathrm{id}) \neq 0$*. Then*

$$\mu_{CZ}(\Psi_S) = \max\{k : \eta_k(S) < 0\} - n. \tag{11.6}$$

Recall that the Conley–Zehnder index as explained in Definition 10.4.1 is only associated to non-degenerate paths of symplectic matrices. Inspired by Theorem 11.2.1 and following [126] we extend the definition of the Conley–Zehnder index to degenerate paths of symplectic matrices, i.e., paths Ψ_S satisfying $\det(\Psi_S(1) - \mathrm{id}) = 0$, in the following way.

Definition 11.2.2. Assume that $\Psi_S \colon [0,1] \to \mathrm{Sp}(n)$ is a degenerate path of symplectic linear maps starting at the identity. Then the *Conley–Zehnder index* of Ψ_S is defined as

$$\mu_{CZ}(\Psi_S) := \max\{k : \eta_k(S) < 0\} - n.$$

In view of Theorem 11.2.1 formula (11.6) is now valid for arbitrary paths of symplectic matrices starting at the identity, regardless if they are degenerate or not. The reason why we extend the Conley–Zehnder index to degenerate paths via the spectral flow formula and not the Maslov index is that via the spectral flow formula the Conley–Zehnder index becomes *lower semi-continuous* whereas the Maslov index is neither lower nor upper semi-continuous.

To prove Theorem 11.2.1 we first show a Lagrangian version of the spectral flow theorem and then use the Lagrangian version to deduce the periodic version, namely Theorem 11.2.1, by looking at its graph. To formulate the Lagrangian version of the spectral flow theorem, we fix a smooth path $S \colon [0,1] \to \mathrm{Sym}(2n)$

of symmetric matrices as before. The Hilbert space we consider consists however not of loops anymore, but of paths satisfying a Lagrangian boundary condition, namely

$$W^{1,2}_{\mathbb{R}^n}([0,1],\mathbb{C}^n) = \{v \in W^{1,2}([0,1],\mathbb{C}^n) : v(0), v(1) \in \mathbb{R}^n\}.$$

We consider the bounded linear operator

$$L_S \colon W^{1,2}_{\mathbb{R}^n}([0,1],\mathbb{C}^n) \to L^2([0,1],\mathbb{C}^n), \quad v \mapsto -J\partial_t v - Sv.$$

This is the same formula as for the operator A_S but note that the domain of the operator now changed. However, thanks to the Lagrangian boundary condition one easily checks that the operator L_S has the same properties as the operator A_S, namely it is a symmetric Fredholm operator of index 0, or considered as an unbounded operator $L_S \colon L^2([0,1],\mathbb{C}^n) \to L^2([0,1],\mathbb{C}^n)$ a self-adjoint operator with dense domain $W^{1,2}_{\mathbb{R}^n}([0,1],\mathbb{C}^n)$. In particular, L_S has the same spectral properties as the operator A_S. The eigenvalues of the operator L_0 are $\pi\ell$ for every $\ell \in \mathbb{Z}$ with corresponding eigenvectors $v(t) = v_0 e^{\pi i \ell t}$ where $v_0 \in \mathbb{R}^n$. In particular, the multiplicity of each eigenvalue is n. Just as in the periodic case we define the section η_k for $k \in \mathbb{Z}$ to the spectral bundle. Just the normalization condition has to be replaced by

(v′) The sections η_k are normalized such that $\eta_j(0) = 0$ for $j \in \{1, \dots, n\}$.

We associate to S the path of Lagrangians

$$\lambda_S \colon [0,1] \to \Lambda(n), \quad \lambda_S(t) = \Psi_S(t)\mathbb{R}^n.$$

We are now in position to state the Lagrangian version of the spectral flow theorem.

Theorem 11.2.3. *Assume that $\Psi_S(1)\mathbb{R}^n \cap \mathbb{R}^n = \{0\}$. Then*

$$\mu_{\mathbb{R}^n}(\lambda_S) = \max\{k : \eta_k(S) < 0\} - \frac{n}{2}.$$

Proof. Recall from (10.1) that the Lagrangian Grassmannian is stratified as

$$\Lambda = \bigcup_{k=0}^{n} \Lambda^k_{\mathbb{R}^n}$$

where $\Lambda^1_{\mathbb{R}^n}$ is the Maslov pseudo-cocycle whose closure is given by

$$\overline{\Lambda^1_{\mathbb{R}^n}} = \bigcup_{k=1}^{n} \Lambda^k_{\mathbb{R}^n}.$$

In particular,

$$\Lambda^0_{\mathbb{R}^n} = \left(\overline{\Lambda^1_{\mathbb{R}^n}}\right)^c$$

is the complement of the closure of the Maslov pseudo-cocycle. Recall further that $\Lambda^0_{\mathbb{R}^n}$ is connected. Indeed, it was shown in Formula (10.6) that $\Lambda^0_{\mathbb{R}^n}$ can be identified with

$$\Lambda^0_{\mathbb{R}^n} = S^2(i\mathbb{R}^n),$$

the vector space of quadratic forms on $i\mathbb{R}^n$, so $\Lambda^0_{\mathbb{R}^n}$ is actually contractible. Note the following equivalences

$$0 \notin \mathfrak{S}(L_S) \Longleftrightarrow \ker L_S = \{0\} \Longleftrightarrow \Psi_S(1)\mathbb{R}^n \cap \mathbb{R}^n = \{0\} \Longleftrightarrow \Psi_S(1)\mathbb{R}^n \in \Lambda^0_{\mathbb{R}^n}. \tag{11.7}$$

For a path $S \in C^\infty([0,1], \mathrm{Sym}(2n))$ define

$$\widetilde{\mu}(S) := \max\{k : \eta_k(S) < 0\}.$$

Consider a homotopy in $\mathcal{P}$. Namely, let $S \in C^\infty([0,1] \times [0,1], \mathrm{Sym}(2n))$ and abbreviate $S_r = S(\cdot, r)$ for the homotopy parameter $r \in [0,1]$. Assume that during the homotopy we never cross the closure of the Maslov pseudo-cocycle, namely

$$\Psi_{S_r}(1)\mathbb{R}^n \in \Lambda^0_{\mathbb{R}^n}, \quad r \in [0,1].$$

We conclude from (11.7) and the continuity of eigenvalues under perturbation that

$$\widetilde{\mu}(S_0) = \widetilde{\mu}(S_1).$$

By homotopy invariance of the intersection number we also have

$$\mu_{\mathbb{R}^n}(\lambda_{S_0}) = \mu_{\mathbb{R}^n}(\lambda_{S_1}).$$

It was shown in Theorem 10.1.1 that the fundamental group of the Lagrangian Grassmannian satisfies $\pi_1(\Lambda) = \mathbb{Z}$ with generator

$$t \mapsto \begin{pmatrix} e^{i\pi t} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \mathbb{R}^n.$$

Because $\Lambda^0_{\mathbb{R}^n}$ is connected, it follows that each non-degenerate path $\lambda_S = \Psi_S \mathbb{R}^n$ is homotopic through a path with endpoints in $\Lambda^0_{\mathbb{R}^n}$ to a path of the form

$$\lambda_k \colon [0,1] \to \mathrm{Sp}(n), \quad t \mapsto \begin{pmatrix} e^{i\pi(\frac{1}{2}+k)t} & & & \\ & e^{i\frac{\pi}{2}t} & & \\ & & \ddots & \\ & & & e^{i\frac{\pi}{2}t} \end{pmatrix} \mathbb{R}^n$$

for some $k \in \mathbb{Z}$. If

$$S_k = \frac{\pi}{2} \begin{pmatrix} 1+2k & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix},$$

then we obtain

$$\lambda_k = \lambda_{S_k}.$$

In view of the homotopy invariance of $\widetilde{\mu}$ and $\mu_{\mathbb{R}^n}$, it suffices to show that

$$\mu_{\mathbb{R}^n}(\lambda_k) = \widetilde{\mu}(S_k) - \frac{n}{2}.$$

But both sides are equal to $\frac{n}{2} + k$ and hence the theorem is proved. $\square$

We now use Theorem 11.2.3 to prove Theorem 11.2.1.

Proof of Theorem 11.2.1. Recall that if (V, ω) is a symplectic vector space, then the diagonal $\Delta = \{(v, v) : v \in V\} \subset V \oplus V$ is a Lagrangian subspace in the symplectic vector space $(V \oplus V, -\omega \oplus \omega)$. If $\Psi\colon [0,1] \to \mathrm{Sp}(V)$ is a smooth path of linear symplectic transformations starting at the identity, i.e., $\Psi(0) = \mathrm{id}$, and Γ_Ψ is the graph of Ψ, then this graph $\Gamma_\Psi\colon [0,1] \to \Lambda(V \oplus V)$ is a path in the Lagrangian Grassmannian with the property that $\Gamma_{\Psi(0)} \in \Delta$. The Conley–Zehnder index of Ψ is by definition

$$\mu_{CZ}(\Psi) = \mu_\Delta(\Gamma_\Psi).$$

Choose a complex structure $J\colon V \to V$, i.e., a linear map satisfying $J^2 = -\mathrm{id}$ which is ω-compatible in the sense that $\omega(\cdot, J\cdot)$ is a scalar product on V. Then $-J \oplus J$ is a complex structure on $V \oplus V$ and

$$-\omega \oplus \omega(\cdot, -J \oplus J\cdot) = \omega(\cdot, J\cdot) \oplus \omega(\cdot, J\cdot)$$

is a scalar product on $V \oplus V$. Define a Hilbert space isomorphism

$$\Gamma\colon W^{1,2}(S^1, V) \to W^{1,2}_\Delta([0,1], V \oplus V)$$

which associates to $v \in W^{1,2}(S^1, V)$ the map

$$\Gamma(v)(t) = \Big(v\big(1 - \tfrac{t}{2}\big), v\big(\tfrac{t}{2}\big)\Big), \quad t \in [0,1].$$

Note that because v was periodic, i.e., $v(0) = v(1)$, it holds that

$$\Gamma(v)(0) = \Big(v(0), v(1)\Big) = \Big(v(0), v(0)\Big) \in \Delta, \quad \Gamma(v)(1) = \Big(v\big(\tfrac{1}{2}\big), v\big(\tfrac{1}{2}\big)\Big) \in \Delta$$

such that $\Gamma(v)$ actually lies in the space $W^{1,2}_\Delta([0,1], V \oplus V)$. Moreover, Γ extends to a map

$$\Gamma\colon L^2(S^1, V) \to L^2([0,1], V \oplus V)$$

by the same formula. With respect to the L^2-inner product Γ is an isometry up to a factor $\sqrt{2}$, indeed

$$\begin{aligned}
||\Gamma(v)||_{L^2} &= \left(\int_0^1 ||\Gamma(v)(t)||^2 dt\right)^{1/2} \\
&= \left(\int_0^1 \left(\left|\left|v\left(1-\tfrac{t}{2}\right)\right|\right|^2 + \left|\left|v\left(\tfrac{t}{2}\right)\right|\right|^2\right) dt\right)^{1/2} \\
&= \left(\int_0^{1/2} \left(\left|\left|v\left(\tfrac{1}{2}+t\right)\right|\right|^2 + \left|\left|v(t)\right|\right|^2\right) 2dt\right)^{1/2} \\
&= \sqrt{2}\left(\int_0^1 ||v(t)||^2 dt\right)^{1/2} \\
&= \sqrt{2}||v||_{L^2}.
\end{aligned}$$

Abbreviate

$$\mathcal{P}(V) = C^\infty([0,1], \mathrm{Sym}(V))$$

where $\mathrm{Sym}(V)$ denotes the vector space of self-adjoint linear maps with respect to the inner product $\omega(\cdot, J\cdot)$. Define a map

$$\Gamma\colon \mathcal{P}(V) \to \mathcal{P}(V \oplus V)$$

which associates to $S \in \mathcal{P}(V)$ the path of self-adjoint linear maps in $V \oplus V$

$$\Gamma(S)(t) = \frac{1}{2}\Big(S\big(1-\tfrac{t}{2}\big), S\big(\tfrac{t}{2}\big)\Big).$$

The operators

$$A_S\colon W^{1,2}(S^1,V) \to L^2(S^1,V), \quad L_{\Gamma(S)}\colon W^{1,2}_\Delta([0,1], V \oplus V) \to L^2([0,1], V \oplus V)$$

are related by

$$\Gamma(A_S v) = L_{\Gamma(S)}\Gamma v, \quad v \in W^{1,2}(S^1,V).$$

Moreover, v is an eigenvector of A_S to the eigenvalue η if and only if $\Gamma(v)$ is an eigenvector of $L_{\Gamma(S)}$ to the eigenvalue $\frac{\eta}{2}$, consistent with the fact that Γ is an isometry up to a factor $\sqrt{2}$ as checked above. In particular, we have

$$\mathfrak{S}(L_{\Gamma(S)}) = \frac{1}{2}\mathfrak{S}(A_S).$$

We conclude that

$$\max\{k : \eta_k(S) < 0\} = \max\{k : \eta_k(\Gamma(S)) < 0\}$$

so that we obtain from Theorem 11.2.3

$$\begin{aligned}
\mu_\Delta(\Psi_{\Gamma(S)}\Delta) &= \max\{k : \eta_k(S) < 0\} - \frac{\dim(V \oplus V)}{2} \\
&= \max\{k : \eta_k(S) < 0\} - \dim V.
\end{aligned} \tag{11.8}$$

Note that the path of symplectomorphisms of $V \oplus V$ generated by $\Gamma(S)$ satisfies

$$\Psi_{\Gamma(S)}(t) = \big(\Psi_S\big(1-\tfrac{t}{2}\big)\Psi_S(1)^{-1}, \Psi_S\big(\tfrac{t}{2}\big)\big), \quad t \in [0,1].$$

Therefore, if we apply this formula to the diagonal, we obtain

$$\Psi_{\Gamma(S)}(t)\Delta = \Gamma_{\Psi_S(1)\Psi_S(1-\frac{t}{2})^{-1}\Psi_S(\frac{t}{2})}.$$

Note that the path of symplectic matrices

$$t \mapsto \Psi_S(1)\Psi_S\big(1-\tfrac{t}{2}\big)^{-1}\Psi_S\big(\tfrac{t}{2}\big), \quad t \in [0,1]$$

is homotopic with fixed endpoints to the path

$$t \mapsto \Psi_S(t), \quad t \in [0,1]$$

via the homotopy

$$(t,r) \mapsto \Psi_S(1)\Psi_S\big(1-\tfrac{t(1-r)}{2}\big)^{-1}\Psi_S\big(\tfrac{t(1+r)}{2}\big), \quad (t,r) \in [0,1]\times[0,1].$$

Consequently, by homotopy invariance of the Maslov index it holds that

$$\mu_\Delta(\Gamma_{\Psi_S}) = \mu_\Delta(\Psi_{\Gamma(S)}\Delta). \tag{11.9}$$

Combining (11.8) and (11.9), we obtain

$$\mu_{CZ}(\Psi_S) = \mu_\Delta(\Gamma_{\Psi_S}) = \max\{k : \eta_k(S) < 0\} - \dim V.$$

This finishes the proof of the theorem. □

11.3 Winding numbers of eigenvalues

The spectral flow theorem is difficult to apply directly since it requires that one knows the numbering of the eigenvalues. But to obtain this numbering one has to understand the bifurcation of the eigenvalues during a homotopy. A very fruitful idea of Hofer, Wysocki and Zehnder [122] is to use the *winding number* to keep track of the numbering of the eigenvalues. This only works if the dimension of the symplectic vector space is two since otherwise the winding number cannot be defined. However, in dimension two this idea led to fantastic applications.

We now restrict our attention to the two-dimensional symplectic vector space $(\mathbb{C}, \omega)$ and consider $S \in C^\infty([0,1], \mathrm{Sym}(2))$ meaning a smooth path of symmetric 2×2-matrices. Recall the operator

$$A_S \colon W^{1,2}(S^1, \mathbb{C}) \to L^2(S^1, \mathbb{C}), \quad v \mapsto -J\partial_t v - S(t)v.$$

Suppose that η is an eigenvalue of A_S and v is a eigenvector of A_S for the eigenvalue η, i.e.,

$$A_S v = \eta v.$$

This means that v is a solution of the linear first-order ODE

$$-J\partial_t v = (S + \eta)v.$$

Since v as an eigenvector cannot vanish identically, it follows from the ODE above that

$$v(t) \neq 0 \in \mathbb{C}, \quad \forall\, t \in S^1.$$

Hence we get a map

$$\gamma_v \colon S^1 \to S^1, \quad t \mapsto \frac{v(t)}{||v(t)||}.$$

We define the *winding number* of the eigenvector v to be

$$w(v) := \deg(\gamma_v) \in \mathbb{Z}$$

where $\deg(\gamma_v)$ denotes the degree of the map γ_v. The following crucial lemma of Hofer, Wysocki and Zehnder appeared in [122].

Lemma 11.3.1. *Assume that v_1 and v_2 are eigenvectors to the* same *eigenvalue η. Then $w(v_1) = w(v_2)$.*

Proof. If v_2 is a scalar multiple of v_1, the lemma is obvious. Hence we can assume that v_1 and v_2 are linearly independent. Define

$$v \colon S^1 \to \mathbb{C}, \quad v(t) = v_1(t)\overline{v_2(t)}.$$

It follows that

$$\deg(\gamma_v) = \deg(\gamma_{v_1}) - \deg(\gamma_{v_2}).$$

It therefore remains to show that

$$\deg(\gamma_v) = 0.$$

This follows from the following claim.

Claim: $v(t) \notin \mathbb{R}$, $\forall\, t \in S^1$.

To prove the claim we argue by contradiction and assume that there exists $t_0 \in S^1$ such that

$$v(t_0) \in \mathbb{R}.$$

Hence there exists $\tau \in \mathbb{R} \setminus \{0\}$ such that

$$v_1(t_0) = \tau v_2(t_0).$$

Define

$$v_3 \colon S^1 \to \mathbb{C}, \quad v_3 = v_1 - \tau v_2.$$

Since $Av_1 = \eta v_1$ and $Av_2 = \eta v_2$, it follows that

$$Av_3 = \eta v_3$$

which implies that v_3 is a solution of a linear first-order ODE. On the other hand

$$v_3(t_0) = v_1(t_0) - \tau v_2(t_0) = 0$$

and therefore

$$v_3(t) = 0, \quad \forall\, t \in S^1.$$

By definition of v_3 this implies that v_1 and v_2 are linearly dependent. This contradiction proves the claim and hence the lemma. □

As a consequence of Lemma 11.3.1 we can now associate to every $\eta \in \mathfrak{S}(A_S)$ a winding number, by

$$w(\eta) := w(\eta, S) := w(v) \tag{11.10}$$

where v is any eigenvector of A_S to the eigenvalue η. Indeed, Lemma 11.3.1 tells us that this is well defined, independent of the choice of the eigenvector. We refer to $w(\eta)$ as the *winding number of the eigenvalue* η.

Let us examine the case $S = 0$. In this case $A_0 = -J\partial_t$, and the eigenvalues are $2\pi\ell$ for every $\ell \in \mathbb{Z}$ with corresponding eigenvectors $v_0 e^{2\pi i \ell t}$ where $v_0 \in \mathbb{C}$. We conclude that

$$w(2\pi\ell, 0) = \ell. \tag{11.11}$$

In particular, we see that in the case $S = 0$ the winding is a monotone function in the eigenvalue. This is true in general.

Corollary 11.3.2 (Monotonicity)**.** *Assume that $S \in C^\infty([0,1], \mathrm{Sym}(2))$. Then the map*

$$w \colon \mathfrak{S}(A_S) \to \mathbb{Z}, \quad \eta \mapsto w(\eta)$$

is monotone, i.e.,

$$\eta \leq \eta' \Longrightarrow w(\eta) \leq w(\eta').$$

Proof. Consider the homotopy $r \mapsto rS$ for $r \in [0,1]$. By (11.11) the assertion is true for $r = 0$. Since the eigenvalues vary continuously under perturbation, we conclude that the assertion of the corollary is true for every $r \in [0,1]$. □

By our convention of numbering the eigenvalues we obtain from (11.11) that

$$w(\eta_{2\ell}) = w(\eta_{2\ell-1}) = \ell - 1, \quad \ell \in \mathbb{Z}. \tag{11.12}$$

Define

$$\alpha(S) := \max\left\{w(\eta, S) : \eta \in \mathfrak{S}(A_S) \cap (-\infty, 0)\right\} \in \mathbb{Z} \tag{11.13}$$

and the *parity*

$$p(S) := \begin{cases} 0 & \text{if } \exists\, \eta \in \mathfrak{S}(A_S) \cap [0,\infty),\ \alpha(S) = w(\eta, S) \\ 1 & \text{else.} \end{cases} \tag{11.14}$$

The spectral flow theorem for two-dimensional symplectic vector spaces gives rise to the following description of the Conley–Zehnder index.

Theorem 11.3.3. *Assume that* $S \in C^\infty([0,1], \mathrm{Sym}(2))$*. Then the Conley–Zehnder index satisfies*

$$\mu_{CZ}(\Psi_S) = 2\alpha(S) + p(S).$$

Proof. In view of Theorem 11.2.1 if Ψ_S is non-degenerate the Conley–Zehnder index is given by

$$\mu_{CZ}(\Psi_S) = \max\{k : \eta_k(S) < 0\} - 1 \tag{11.15}$$

and if Ψ_S is degenerate we use this formula as definition of the Conley–Zehnder index. By (11.12) we have

$$\alpha(S) = w(\eta_{2\alpha(S)+1}) = w(\eta_{2\alpha(S)+2}).$$

Therefore

$$\{\eta \in \mathfrak{S}(A_S) : w(\eta) = \alpha(S)\} = \{\eta_{2\alpha(S)+1}, \eta_{2\alpha(S)+2}\}.$$

Since $\eta_k \leq \eta_{k+1}$, the definition of $\alpha(S)$ implies that

$$\eta_{2\alpha(S)+1} < 0, \quad \eta_{2\alpha(S)+3} \geq 0.$$

Therefore, by definition of the parity we get

$$p(S) = \begin{cases} 0 & \text{if } \eta_{2\alpha(S)+2} \geq 0 \\ 1 & \text{if } \eta_{2\alpha(S)+2} < 0. \end{cases}$$

Hence

$$\max\{k : \eta_k(S) < 0\} = 2\alpha(S) + 1 + p(S)$$

and the theorem follows from (11.15). □

Chapter 12

Convexity

12.1 Convex hypersurfaces

We first recall some standard facts about convex hypersurfaces in the Euclidean vector space following [151, Chapter 7.5], [234, Chapter 13]. Assume that $S \subset \mathbb{R}^{n+1}$ is a closed, connected hypersurface. It follows that we get a decomposition

$$\mathbb{R}^{n+1} \setminus S = M_- \cup M_+$$

into two connected components, where M_- is bounded and M_+ is unbounded. If $p \in S$ we denote by $N(p)$ the unit normal vector of S pointing into the unbounded component M_+. This leads to a smooth map

$$N\colon S \to S^n$$

referred to as the *Gauss map* which defines an orientation on S. Because $T_{N(p)}S^n$ and T_pS are parallel planes we can identify them canonically so that the differential of the Gauss map becomes a linear map

$$dN(p)\colon T_pS \to T_pS.$$

The *Gauss–Kronecker curvature* at $p \in S$ is defined as

$$K(p) := \det dN(p).$$

Note that if $n = 2$, i.e., S is a two-dimensional surface, the Gauss–Kronecker curvature coincides with the Gauss curvature of the surface, which is intrinsic.

We write S as a level set, i.e., we pick $f \in C^\infty(\mathbb{R}^{n+1}, \mathbb{R})$ such that 0 is a regular value of f and

$$S = f^{-1}(0).$$

We choose f in such a way that

$$M_- = \{x \in \mathbb{R}^{n+1} : f(x) < 0\}, \quad M_+ = \{x \in \mathbb{R}^{n+1} : f(x) > 0\}.$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_12

It follows that

$$N = \frac{\nabla f}{||\nabla f||}.$$

We abbreviate by

$$H_f(p)\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$$

the Hessian of f at p. We get for a point $p \in S$ and two tangent vectors $v, w \in T_pS = \nabla f(p)^\perp$ the equality

$$\langle v, dN(p)w\rangle = \frac{1}{||\nabla f(p)||}\langle v, H_f(p)w\rangle.$$

In particular, if

$$\Pi_p\colon \mathbb{R}^{n+1} \to T_pS$$

is the orthogonal projection, we can write

$$dN(p) = \frac{1}{||\nabla f(p)||}\Pi_p H_f(p)|_{T_pS} = \frac{1}{||\nabla f(p)||}\Pi_p H_f(p)\Pi_p^*$$

implying that $dN(p)$ is self-adjoint. We make the following definition

Definition 12.1.1. The hypersurface S is called *strictly convex*, if $K(p) > 0$ for every $p \in S$.

If the hypersurface is written as the level set of a function $S = f^{-1}(0)$, then for practical purposes it is useful to note that strict convexity is equivalent to the assertion that

$$\det(\Pi_p H_f(p)|_{\nabla f(p)^\perp}) > 0, \quad \forall\, p \in S.$$

However, observe that the notion of convexity only depends on S and not on the choice of the function f.

Lemma 12.1.2. *The hypersurface S is strictly convex if and only if $dN(p)$ is positive definite for every $p \in S$.*

Proof. The implication "$\Longleftarrow$" is obvious. It remains to check the implication "$\Longrightarrow$". Because $dN(p)$ is self-adjoint, it follows that $dN(p)$ has n real eigenvalues counted with multiplicity. If $\mathfrak{S}(dN(p)) \subset \mathbb{R}$ is the spectrum of $dN(p)$, then define

$$k\colon S \to \mathbb{R}, \quad p \mapsto \min\{r : r \in \mathfrak{S}(dN(p))\}$$

as the smallest eigenvalue of $dN(p)$. The function k is a continuous function on S and we claim

$$k(p) > 0, \quad \forall\, p \in S. \tag{12.1}$$

We prove (12.1) in two steps. We first check

Step 1: *There exists $p_0 \in S$ such that $k(p_0) > 0$.*

To prove the assertion of Step 1 we denote the closed ball of radius r with $r \in (0, \infty)$ by $D(r) = \{x \in \mathbb{R}^{n+1} : ||x|| \leq r\}$ and set

$$r_S := \min\{r \in (0, \infty) : S \subset D(r)\}.$$

Because S is compact, the radius r_S is finite. Moreover, there exists $p_0 \in \partial D_{r_S} = S^n_{r_S} = \{x \in \mathbb{R}^{n+1} : ||x|| = r_S\}$, the sphere of radius r_S, such that $p_0 \in S$ as well. At this point $S^n_{r_S}$ touches S from outside so that we obtain

$$k^S(p_0) \geq k^{S^n_{r_S}}(p_0) = \frac{1}{r_S} > 0.$$

This finishes the proof of Step 1.

Step 2: *We prove* (12.1).

Because $K(p) = \det dN(p) > 0$ for every $p \in S$, it follows that $k(p) \neq 0$ for every $p \in S$. Since S is connected and the function k is continuous, we either have $k(p) > 0$ for every $p \in S$ or $k(p) < 0$ for every $p \in S$. By Step 1 we conclude that $k(p) > 0$ for every $p \in S$. This establishes (12.1).

We are now in position to prove the lemma. Because $k(p) > 0$, it follows that

$$\mathfrak{S}(dN(p)) \in (0, \infty) \text{ for all } p \in S.$$

Therefore $dN(p)$ is positive definite for every $p \in S$ and the lemma is proved. □

For the following lemma recall that $M_- \subset \mathbb{R}^{n+1}$ denotes the bounded region of $\mathbb{R}^{n+1} \setminus S$.

Lemma 12.1.3. *Assume that S is strictly convex. Then M_- is convex in the sense that if $x, y \in M_-$ and $\lambda \in [0, 1]$ then $\lambda x + (1 - \lambda y) \in M_-$, i.e., the line segment between x and y is contained in M_-.*

Proof. The assertion of the lemma is true in any dimension. However, the proof we present just works if $n \geq 2$.

For $p \in S$ we introduce the half-space

$$H_p := \{x \in \mathbb{R}^{n+1} : \langle p - x, N(p) \rangle > 0\}.$$

We claim

$$M_- = \bigcap_{p \in S} H_p. \tag{12.2}$$

We first check

$$\bigcap_{p \in S} H_p \subset M_-. \tag{12.3}$$

To prove that, we note the following equivalences

$$\begin{aligned}\bigcap_{p\in S} H_p \subset M_- &\iff M_+ \cup S = M_-^c \subset \Big(\bigcap_{p\in S} H_p\Big)^c = \bigcup_{p\in S} H_p^c \\ &\iff M_+ = \bigcup_{p\in S} \big\{x \in \mathbb{R}^{n+1} : \langle x-p, N(p)\rangle > 0\big\}.\end{aligned}$$

If $x \in M_+$, choose $p_0 \in S$ and consider the line segment

$$[0,1] \to \mathbb{R}^{n+1}, \quad t \mapsto (1-t)p_0 + tx.$$

Define

$$t_0 := \max\big\{t \in [0,1] : (1-t)p_0 + tx \in S\big\}.$$

Because $x \in M_+$, it follows that

$$t_0 < 1.$$

We set

$$p := (1-t_0)p_0 + t_0 x \in S.$$

Since $x \in M_+$, it follows that

$$\langle x-p, N(p)\rangle > 0.$$

This establishes (12.3). Note that for the proof of (12.3) we did not use yet the convexity of S. We next check that

$$M_- \subset \bigcap_{p\in S} H_p. \tag{12.4}$$

We need to show that for every $p \in S$ it holds that

$$M_- \subset H_p.$$

For $u \in S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$ consider

$$F_u \colon \mathbb{R}^{n+1} \to \mathbb{R}, \quad x \mapsto \langle x, u\rangle$$

and abbreviate

$$f_u := F_u|_S \colon S \to \mathbb{R}$$

the restriction of F_u to S. We next discuss the critical points of f_u. Recall from the method of Lagrange multipliers that if $F \colon \mathbb{R}^{n+1} \to \mathbb{R}$ is a smooth function and $f = F|_S$, then $p \in \mathrm{crit} f$ if and only if there exists $\lambda_p \in \mathbb{R}$, referred to as the Lagrange multiplier, such that

$$\nabla f(p) = \lambda_p N(p).$$

Moreover, if $\Pi_p \colon \mathbb{R}^{n+1} \to T_pS$ denotes the orthogonal projection, the Hessian of f at p is given by

$$H_f(p) = \big(\Pi_p H_F(p)|_{T_pS} + \lambda_p dN(p)\big) \colon T_pS \to T_pS.$$

In the case that interests us, we have

$$\nabla F_u(p) = u, \quad H_{F_u}(p) = 0$$

and therefore $p \in \mathrm{crit} f_u$ if and only if

$$u = \lambda_p N(p)$$

where

$$\lambda_p = \pm 1$$

because $||u|| = ||N(p)|| = 1$. Moreover, the Hessian is given by

$$H_{f_u}(p) = \lambda_p dN(p).$$

Because S is strictly convex, Lemma 12.1.2 tells us that $dN(p)$ is positive definite for every $p \in S$. Therefore H_{f_u} is either positive definite or negative definite for every $p \in S$. We have shown that $f_u \colon S \to \mathbb{R}$ is a Morse function all whose critical points are either maxima or minima. At this point we need the assumption that $n \geq 2$, i.e., that the dimension of S is at least two. Namely, because S is connected, it follows that in this case f_u has precisely one global maximum and precisely one global minimum and no other critical points.

For $p \in S$ choose $u = N(p)$. It follows that

$$p \in \mathrm{crit} f_{N(p)}$$

with

$$\lambda_p = 1.$$

Therefore p is the unique global maximum of $f_{N(p)}$. In particular, for every $q \in S \setminus \{p\}$ we have

$$\langle q, N(p)\rangle = f_{N(p)}(q) < f_{N(p)}(p) = \langle p, N(p)\rangle$$

implying that

$$0 < \langle p, N(p)\rangle - \langle q, N(p)\rangle = \langle p - q, N(p)\rangle, \quad \forall\, q \in S \setminus \{p\}.$$

It follows that

$$S \setminus \{p\} \subset H_p$$

and consequently

$$M_- \subset H_p.$$

This establishes (12.4) and therefore, together with (12.3), we obtain (12.2).

In view of (12.2) the lemma can now be proved as follows. Note that H_p is convex for every $p \in S$. Because the intersection of convex sets is convex, it follows that $\bigcap_{p \in S} H_p$ is convex as well. Hence (12.2) implies the lemma. $\square$

12.2 Convexity implies dynamical convexity

Recall that a starshaped hypersurface in $\mathbb{C}^2$ is called dynamically convex if the Conley–Zehnder indices of all its periodic orbits are at least three. A strictly convex hypersurface in $\mathbb{C}^2$ is starshaped and we prove in this section that it is dynamically convex, explaining the notion "dynamically convex".

Suppose that $\Sigma \subset \mathbb{C}^2$ is strictly convex. Because convexity is preserved under affine transformations we can assume without loss of generality that $0 \in M_-$, the bounded part of $\mathbb{C}^2 \setminus \Sigma$. Therefore M_- is star-shaped by Lemma 12.1.3. It follows that the restriction $\lambda|_\Sigma$ of the standard Liouville form

$$\lambda = \frac{1}{2}(x_1 dy_1 - y_1 dx_1 + x_2 dy_2 - y_2 dx_2)$$

to Σ is a contact form on Σ.

The following theorem appeared in [126]

Theorem 12.2.1 (Hofer–Wysocki–Zehnder). *Assume that $\Sigma \subset \mathbb{C}^2$ is a strictly convex hypersurface such that 0 lies in the bounded part of $\mathbb{C}^2 \setminus \Sigma$. Then the contact manifold $(\Sigma, \lambda|_\Sigma)$ is dynamically convex where λ is the standard Liouville form on $\mathbb{C}^2$.*

As preparation for the proof suppose that $\Sigma \subset \mathbb{C}^2$ meets the assumption of Theorem 12.2.1. For $z \in \mathbb{C}^2 \setminus \{0\}$ there exists a unique $r_z \in (0, \infty)$ with the property that

$$r_z z \in \Sigma.$$

Define

$$F_\Sigma \colon \mathbb{C}^2 \setminus \{0\} \to \mathbb{R}, \quad z \mapsto \frac{1}{r_z^2}.$$

It follows that

$$\Sigma = F_\Sigma^{-1}(1).$$

We denote by X_{F_Σ} the Hamiltonian vector field of F_Σ with respect to the standard symplectic structure $\omega = d\lambda$ on $\mathbb{C}^2$ defined implicitly by the condition

$$dF_\Sigma = \omega(\cdot, X_{F_\Sigma}).$$

Lemma 12.2.2. *For $z \in \Sigma$ it holds that*

$$X_{F_\Sigma}(z) = R(z)$$

where R is the Reeb vector field of (Σ, λ_Σ).

Proof. For $v \in T_z\Sigma$ we compute

$$d\lambda(z)(X_{F_\Sigma}(z), v) = \omega_z(X_{F_\Sigma}, v) = -dF_\Sigma(z)v = 0$$

where for the last equality we have used that $\Sigma = F_\Sigma^{-1}(1)$. It follows that

$$X_{F_\Sigma}|_\Sigma \in \ker d\lambda|_\Sigma.$$

It remains to show that

$$\lambda(X_{F_\Sigma})|_\Sigma = 1.$$

Note that F_Σ is homogeneous of degree 2, i.e.,

$$F_\Sigma(rz) = r^2 F_\Sigma(z), \quad z \in \mathbb{C}^2 \setminus \{0\},\ r > 0.$$

Consequently

$$dF_\Sigma(z)z = \frac{d}{dr}\bigg|_{r=1} F_\Sigma(rz) = 2rF_\Sigma(z)\big|_{r=1} = 2F_\Sigma(z) \tag{12.5}$$

which is known as Euler's formula. Note further that if $z, \widehat{z} \in \mathbb{C}$ the equation

$$\lambda(z)\widehat{z} = \frac{1}{2}\omega(z, \widehat{z}) \tag{12.6}$$

holds true. Using (12.5) and (12.6) we compute for $z \in \Sigma = F_\Sigma^{-1}(1)$

$$\lambda(z)(X_{F_\Sigma(z)}) = \frac{1}{2}\omega(z, X_{F_\Sigma}(z)) = \frac{1}{2}dF_\Sigma(z)z = F_\Sigma(z) = 1.$$

This finishes the proof of the lemma. $\square$

Suppose that $\gamma \in C^\infty(S^1, \Sigma)$ is a periodic Reeb orbit of period τ. Abbreviate by

$$\phi_\Sigma^t \colon \mathbb{C}^2 \to \mathbb{C}^2, \quad t \in \mathbb{R}$$

the flow of the vector field X_{F_Σ}. Because γ is a periodic Reeb orbit, we have in view of Lemma 12.2.2 for every $t \in \mathbb{R}$ the equality

$$\phi_\Sigma^{\tau t}(\gamma(0)) = \gamma(t).$$

For $t \in [0,1]$ consider the smooth path of symplectic matrices

$$\Psi_\gamma^t := d\phi_\Sigma^{\tau t}(\gamma(0)) \colon \mathbb{C}^2 \to \mathbb{C}^2. \tag{12.7}$$

Lemma 12.2.3. *The path Ψ_γ^t has the following properties*

(i) $\Psi_\gamma^0 = \mathrm{id}$,

(ii) Ψ_γ^t *satisfies the ODE*

$$\partial_t \Psi_\gamma^t = J\tau H_{F_\Sigma}(\gamma(t))\Psi_\gamma^t,$$

where $J \colon \mathbb{C}^2 \to \mathbb{C}^2$ is multiplication by i and H_{F_Σ} is the Hessian of F_Σ which is positive definite by Lemma 12.1.2,

(iii) $\Psi_\gamma^1\big(R(\gamma(0))\big) = R(\gamma(0))$ *and* $\Psi_\gamma^1\big(\gamma(0)\big) = \gamma(0)$.

Proof. Properties (i) and (ii) are immediate. To explain why the first equation of property (iii) holds, note that because the Hamiltonian vector field X_{F_Σ} is autonomous (time-independent), its flow satisfies

$$\phi_\Sigma^{s+t} = \phi_\Sigma^t \phi_\Sigma^s$$

and therefore

$$X_{F_\Sigma} \circ \phi_\Sigma^t = \frac{d}{ds}\bigg|_{s=0} \phi_\Sigma^{s+t} = \frac{d}{ds}\bigg|_{s=0} \phi_\Sigma^t \phi_\Sigma^s = d\phi_\Sigma^t(X_{F_\Sigma}).$$

Because γ is a periodic Reeb orbit of period τ, we have

$$\phi_\Sigma^\tau(\gamma(0)) = \gamma(0)$$

and therefore

$$R(\gamma(0)) = X_{F_\Sigma}(\gamma(0)) = d\phi_\Sigma^\tau(\gamma(0))\big(X_{F_\Sigma}(\gamma(0))\big) = \Psi_\gamma^1\big(R(\gamma(0))\big).$$

This explains the first equation in property (iii).

It remains to check the second equation in property (iii). We first recall that because F_Σ is homogeneous of degree 2, Euler's formula (12.5) holds for every $z \in \mathbb{C}$. Differentiating once more we obtain

$$d^2F_\Sigma(z)z + dF_\Sigma(z) = 2dF_\Sigma(z),$$

which implies that

$$d^2F_\Sigma(z)z = dF_\Sigma(z).$$

In particular, we have

$$H_{F_\Sigma}(z)z = \nabla F_\Sigma(z)$$

and therefore

$$JH_{F_\Sigma}(z)z = J\nabla F_\Sigma(z) = X_{F_\Sigma}(z). \tag{12.8}$$

We claim that

$$d\phi_\Sigma^t(z)z = \phi_\Sigma^t(z) \in \mathbb{C}^2, \quad \forall\, z \in \mathbb{C}^2. \tag{12.9}$$

To check this equation we fix $z \in \mathbb{C}^2$ and consider the path

$$w\colon \mathbb{R} \to \mathbb{C}^2, \quad t \mapsto d\phi_\Sigma^t(z)z - \phi_\Sigma^t(z).$$

Note that

$$w(0) = d\phi_\Sigma^0(z)z - \phi_\Sigma^0(z) = z - z = 0. \tag{12.10}$$

Moreover,

$$\begin{aligned}
\frac{d}{dt}w(t) &= \frac{d}{dt}d\phi_\Sigma^t(z)z - \frac{d}{dt}\phi_\Sigma^t(z) \qquad (12.11)\\
&= JH_{F_\Sigma}(\phi_\Sigma^t(z))d\phi_\Sigma^t(z)z - X_{F_\Sigma}(\phi_\Sigma^t(z))\\
&= JH_{F_\Sigma}(\phi_\Sigma^t(z))d\phi_\Sigma^t(z)z - JH_{F_\Sigma}(\phi_\Sigma^t(z))\phi_\Sigma^t(z)\\
&= JH_{F_\Sigma}(\phi_\Sigma^t(z))(d\phi_\Sigma^t(z)z - \phi_\Sigma^t(z))\\
&= JH_{F_\Sigma}(\phi_\Sigma^t(z))w(t).
\end{aligned}$$

Here we have used (12.8) in the third equation. Combining (12.10) and (12.11), we obtain

$$w(t) = 0, \quad \forall\, t \in \mathbb{R}.$$

This proves (12.9). Setting $t = \tau$ and $z = \gamma(0)$ we obtain

$$\Psi^1_\gamma\big(\gamma(0)\big) = d\phi^\tau_\Sigma(\gamma(0))\big(\gamma(0)\big) = \phi^\tau_\Sigma(\gamma(0)) = \gamma(0).$$

This finishes the proof of the lemma. □

For the next lemma recall from Lemma 10.2.3 that if $L \in \Lambda$ is the Lagrangian Grassmannian we have a canonical identification

$$T_L\lambda \to S^2(L), \quad \widehat{L} \mapsto Q^{\widehat{L}}$$

where $S^2(L)$ is the space of quadratic forms on L. Recall that if $\Psi \in \mathrm{Sp}(n)$ is a symplectic transformation, then its graph Γ_Ψ is a Lagrangian in the symplectic vector space $(\mathbb{C}^n \oplus \mathbb{C}^n, -\omega \oplus \omega)$. We can now state the next lemma.

Lemma 12.2.4. *Suppose that* $\Psi\colon (-\epsilon, \epsilon) \to \mathrm{Sp}(n)$ *is a smooth path of symplectic matrices. Then for*

$$\Gamma_{\Psi'(0)} = \frac{d}{dt}\bigg|_{t=0} \Gamma_{\Psi(t)} \in T_{\Gamma_{\Psi(0)}}\Lambda(\mathbb{C}^n \oplus \mathbb{C}^n, -\omega \oplus \omega)$$

the corresponding quadratic form

$$Q^{\Gamma_{\Psi'(0)}} \in S^2(\Gamma_{\Psi(0)})$$

is given for $(z, \Psi(0)z) \in \Gamma_{\Psi(0)}$ *where* $z \in \mathbb{C}^n$ *by*

$$Q^{\Gamma_{\Psi'(0)}}(z, \Psi(0)z) = \langle \Psi(0)z, S\Psi_0 z\rangle$$

where

$$S = -J\Psi'(0)\Psi(0)^{-1}.$$

Proof. We choose as Lagrangian complement of $\Gamma_{\Psi(0)}$ the Lagrangian

$$\Gamma_{-\Psi(0)} = \big\{(z, -\Psi(0)z) : z \in \mathbb{C}^n\big\}.$$

If $z \in \mathbb{C}^n$ and $t \in (-\epsilon, \epsilon)$, we define $w_z(t) \in \mathbb{C}^n$ by the condition that

$$\big(z, \Psi(0)z\big) + \big(w_z(t), -\Psi(0)w_z(t)\big) \in \Gamma_{\Psi(t)}$$

or equivalently

$$\big(z + w_z(t), \Psi(0)z - \Psi(0)w_z(t)\big) \in \Gamma_{\Psi(t)}.$$

This implies that

$$\Psi(t)z + \Psi(t)w_z(t) = \Psi(t)(z + w_z(t)) = \Psi(0)z - \Psi(0)w_z(t).$$

Differentiating this expression we get

$$\Psi'(t)z + \Psi'(t)w_z(t) + \Psi(t)w_z'(t) = -\Psi(0)w_z'(t).$$

Because $w_z(0) = 0$ we obtain from that

$$\Psi'(0)z + \Psi(0)w_z'(0) = -\Psi(0)w_z'(0)$$

implying

$$w_z'(0) = -\frac{1}{2}\Psi(0)^{-1}\Psi'(0)z.$$

By definition of the quadratic form we compute, taking advantage of the fact that $\Psi(0)$ is symplectic, that

$$\begin{aligned}
Q^{\Gamma_{\Psi'(0)}}\big(z, \Psi(0)z\big) &= -\omega \oplus \omega\Big((z, \Psi(0)z), (w_z'(0), -\Psi(0)w_z'(0)\Big) \\
&= -\omega(z, w_z'(0)) - \omega(\Psi(0)z, \Psi(0)w_z'(0)) \\
&= -\omega(z, w_z'(0)) - \omega(z, w_z'(0)) \\
&= -2\omega(z, w_z'(0)) \\
&= \omega(z, \Psi(0)^{-1}\Psi'(0)z) \\
&= \omega(\Psi(0)z, \Psi'(0)z) \\
&= \omega(\Psi(0)z, JS\Psi(0)z) \\
&= \langle \Psi(0)z, S\Psi(0)z\rangle.
\end{aligned}$$

This finishes the proof of the lemma. $\square$

Corollary 12.2.5. *Assume that $\Psi\colon [0,1] \to Sp(n)$ is a smooth path of symplectic matrices satisfying $\Psi'(t) = JS(t)\Psi(t)$ with $S(t)$ positive definite for every $t \in [0,1]$. Then the crossing form $C(\Gamma_\Psi, \Delta, t)$ is positive definite for every $t \in [0,1]$.*

Proof. By definition of the crossing form we have

$$C(\Gamma_\Psi, \Delta, t) = Q^{\Gamma_{\Psi'(t)}}|_{\Delta \cap \Gamma_{\Psi(t)}}.$$

The corollary follows now from Lemma 12.2.4 and the assumption that S is positive definite. $\square$

Corollary 12.2.6. *Assume that $\Sigma \subset \mathbb{C}^2$ is a strictly convex hypersurface and γ is a periodic Reeb orbit of period τ on Σ. Then*

$$\mu_\Delta(\Gamma_{\Psi_\gamma}) \geq 3 + \frac{1}{2}\dim\ker\big(d^\xi\phi_R^\tau(\gamma(0)) - \mathrm{id}\big) \geq 3.$$

Proof. Using the definition of the Maslov index for paths (10.13), Corollary 12.2.5 and assertions (i) and (iii) in Lemma 12.2.3 we estimate

$$\begin{aligned}
\mu_\Delta(\Gamma_{\Psi_\gamma}) &= \frac{1}{2}\mathrm{sign}C(\Gamma_{\Psi_\gamma},\Delta,0) + \sum_{0<t<1}\mathrm{sign}C(\Gamma_{\Psi_\gamma},\Delta,t) + \frac{1}{2}\mathrm{sign}C(\Gamma_{\Psi_\gamma},\Delta,1)\\
&= \frac{1}{2}\dim\big(\Delta\cap\Gamma_{\Psi_\gamma}(0)\big) + \sum_{0<t<1}\dim\big(\Delta\cap\Gamma_{\Psi_\gamma}(t)\big)\\
&\quad + \frac{1}{2}\dim\big(\Delta\cap\Gamma_{\Psi_\gamma}(1)\big)\\
&= \frac{1}{2}\dim\big(\Delta\big) + \sum_{0<t<1}\dim\big(\Delta\cap\Gamma_{\Psi_\gamma}(t)\big)\\
&\quad + \frac{1}{2}\Big(2+\dim\ker\big(d^\xi\phi_R^\tau(\gamma(0))-\mathrm{id}\big)\Big)\\
&= 3 + \sum_{0<t<1}\dim\big(\Delta\cap\Gamma_{\Psi_\gamma}(t)\big) + \frac{1}{2}\dim\ker\big(d^\xi\phi_R^\tau(\gamma(0))-\mathrm{id}\big)\\
&\geq 3 + \frac{1}{2}\dim\ker\big(d^\xi\phi_R^\tau(\gamma(0))-\mathrm{id}\big).
\end{aligned}$$

This proves the corollary. $\square$

Proof of Theorem 12.2.1. It follows from Lemma 12.1.3 that Σ bounds a star-shaped domain. Therefore Σ is diffeomorphic to the sphere S^3. In particular, $\pi_2(\Sigma)=\{0\}$ and therefore the homomorphism $I_{c_1}\colon \pi_2(\Sigma)\to\mathbb{Z}$ vanishes for trivial reasons. Assume that γ is a periodic orbit on Σ of period τ. Because Σ is simply connected, γ is contractible so that we can choose a filling disk for γ, i.e., a smooth map $\overline{\gamma}\colon D\to\Sigma\subset\mathbb{C}^2$ satisfying $\overline{\gamma}(e^{2\pi it})=\gamma(t)$ for every $t\in S^1$. Note that the vector bundle $\overline{\gamma}^*\mathbb{C}^2\to D$ splits as

$$\overline{\gamma}^*\mathbb{C}^2 = \overline{\gamma}^*\xi\oplus\overline{\gamma}^*\eta, \tag{12.12}$$

where we abbreviate

$$\eta = \langle X, R\rangle,$$

where the vector field X is defined as

$$X(x)=x,\quad x\in\mathbb{C}^2.$$

Note that we have a canonical trivialization

$$\mathfrak{T}_\eta\colon\overline{\gamma}^*\eta\to D\times\mathbb{C},\quad aX+bR\mapsto a+ib,\ a,b\in\mathbb{R}.$$

It follows from Lemma 12.2.2 and Lemma 12.2.3 that the map $\Psi_\gamma^t\colon\mathbb{C}^2\to\mathbb{C}^2$ for $t\in[0,1]$ respects the splitting (12.12), i.e.,

$$\Psi_\gamma^t\colon\xi_{\gamma(0)}\to\xi_{\gamma(t)},\quad \Psi_\gamma^t\colon\eta_{\gamma(0)}\to\eta_{\gamma(t)}.$$

Indeed, we have
$$\Psi_\gamma^t|_\xi = d^\xi\phi_R^{t\tau}(\gamma(0))$$
and
$$\mathfrak{T}_{\eta,\gamma(t)}\Psi_\gamma^t\mathfrak{T}_{\eta,\gamma(0)}^{-1} = \mathrm{id}\colon \mathbb{C}\to\mathbb{C}.$$
Choose in addition a symplectic trivialization
$$\mathfrak{T}_\xi\colon \overline{\gamma}^*\xi \to D\times\mathbb{C}.$$
Hence we obtain a symplectic bundle map
$$\mathfrak{T} := \mathfrak{T}_\xi \oplus \mathfrak{T}_\eta\colon D\times\mathbb{C}^2 \to D\times\mathbb{C}^2.$$
Because the disk D is contractible, the map $\mathfrak{T}$ is homotopic as a bundle map to the identity map from $D\times\mathbb{C}^2$ to itself. In particular, if we introduce the path of symplectic matrices
$$\Psi_\gamma^{\mathfrak{T}}\colon [0,1]\to \mathrm{Sp}(2)$$
defined for $t\in[0,1]$ as
$$(\Psi_\gamma^{\mathfrak{T}})^t = \mathfrak{T}_{\gamma(t)}\Psi_\gamma^t\mathfrak{T}_{\gamma(0)}^{-1}\colon \mathbb{C}^2\to\mathbb{C}^2$$
we obtain by homotopy invariance of the Maslov index and Corollary 12.2.6
$$\mu_\Delta(\Gamma_{\Psi_\gamma^{\mathfrak{T}}}) = \mu_\Delta(\Gamma_{\Psi_\gamma}) \geq 3 + \frac{1}{2}\dim\ker\big(d^\xi\phi_R^\tau(\gamma(0)) - \mathrm{id}\big). \tag{12.13}$$
If (V_1,ω_1) and (V_2,ω_2) are two symplectic vector spaces, $L_1\subset V_1$, $L_2\subset V_2$ are Lagrangian subspaces, $\lambda_1\colon[0,1]\to\Lambda(V_1,\omega_1)$ is a path of Lagrangians in V_1, and $\lambda_2\colon[0,1]\to\Lambda(V_2,\omega_2)$ is a path of Lagrangians in V_2, then we obtain a Lagrangian
$$L_1\oplus L_2\subset V_1\times V_2$$
and a path of Lagrangians
$$\lambda_1\oplus\lambda_2\colon[0,1]\to\Lambda(V_1\oplus V_2,\omega_1\oplus\omega_2).$$
The Maslov index satisfies
$$\mu_{L_1\oplus L_2}(\lambda_1\oplus\lambda_2) = \mu_{L_1}(\lambda_1)+\mu_{L_2}(\lambda_2).$$
In view of the splitting
$$\Psi_\gamma^{\mathfrak{T}} = \Psi_\gamma^{\mathfrak{T}}|_\xi \oplus \Psi_\gamma^{\mathfrak{T}}|_\eta$$
we obtain
$$\mu_\Delta(\Gamma_{\Psi_\gamma^{\mathfrak{T}}}) = \mu_\Delta(\Gamma_{\Psi_\gamma^{\mathfrak{T}}|_\xi}) + \mu_\Delta(\Gamma_{\Psi_\gamma^{\mathfrak{T}}|_\eta}). \tag{12.14}$$
Because $\Psi_\gamma^{\mathfrak{T}}|_\eta(t) = \mathrm{id}$ for every $t\in[0,1]$ we conclude that
$$\mu_\Delta(\Gamma_{\Psi_\gamma^{\mathfrak{T}}|_\eta}) = 0. \tag{12.15}$$

Combining (12.13), (12.14), and (12.15) we obtain the inequality

$$\mu_\Delta(\Gamma_{\Psi^{\mathfrak{T}}_\gamma|_\xi}) \geq 3 + \frac{1}{2}\dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big). \tag{12.16}$$

We distinguish two cases

Case 1: *The periodic orbit γ is non-degenerate.*

In this case it follows from the definition of the Conley–Zehnder index that

$$\mu_{CZ}(\gamma) = \mu_\Delta(\Gamma_{\Psi^{\mathfrak{T}}_\gamma|_\xi}).$$

It now follows from (12.16) that

$$\mu_{CZ}(\gamma) \geq 3.$$

Case 2: *The periodic orbit γ is degenerate.*

In this case we consider for $\epsilon > 0$ a smooth path of symplectic matrices $\Psi_\epsilon \colon [0, 1+\epsilon] \to \mathrm{Sp}(1)$ with the property that $\Psi_\epsilon(t) = \Psi^{\mathfrak{T}}_\gamma|_\xi(t)$ for every $t \in [0,1]$ and there are no further crossings of Γ_{Ψ_ϵ} with the closure of the Maslov pseudo-cocycle $\overline{\Lambda^1_\Delta}$ in $(1, 1+\epsilon]$. It follows from the definition of the Maslov index for paths that

$$\begin{aligned}
\mu_\Delta(\Gamma_{\Psi_\epsilon}) &= \mu_\Delta(\Gamma_{\Psi^{\mathfrak{T}}_\gamma|_\xi}) + \frac{1}{2}\mathrm{sign}C(\Gamma_{\Psi^{\mathfrak{T}}_\gamma|_\xi}, \Delta, 1) \qquad (12.17)\\
&= \mu_\Delta(\Gamma_{\Psi^{\mathfrak{T}}_\gamma|_\xi}) + \frac{1}{2}\dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&\geq 3 + \frac{1}{2}\dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&\quad + \frac{1}{2}\dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&= 3 + \dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big).
\end{aligned}$$

Here we have used in the second equation Lemma 12.2.4 together with the assumption that Σ is strictly convex and for the inequality we used (12.16). Because the path Γ_{Ψ_ϵ} is non-degenerate, we have by definition of the Conley–Zehnder index

$$\mu_\Delta(\Gamma_{\Psi_\epsilon}) = \mu_{CZ}(\Psi_\epsilon). \tag{12.18}$$

In view of the continuity of eigenvalues we have

$$\begin{aligned}
\mu_{CZ}(\Psi^{\mathfrak{T}}_\gamma) &\geq \lim_{\epsilon\to 0}\mu_{CZ}(\Psi_\epsilon) - \dim\ker\big(A_{\Psi^{\mathfrak{T}}_\gamma}\big) \qquad (12.19)\\
&= \mu_{CZ}(\Psi_\epsilon) - \dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&\geq 3 + \dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&\quad - \dim\ker\big(d^\xi\phi^\tau_R(\gamma(0)) - \mathrm{id}\big)\\
&= 3,
\end{aligned}$$

where for the second inequality we have used (12.17) and (12.18). Because the Conley–Zehnder index satisfies $\mu_{CZ}(\gamma) = \mu_{CZ}(\Psi^{\mathfrak{T}}_\gamma)$, inequality (12.19) finishes the proof of the theorem. □

12.3 Hamiltonian flow near a critical point of index 1

In this section we will see how a connected sum can give a dynamical obstruction to convexity. Consider a symplectic manifold (M^4, ω) with Hamiltonian $H : M \to \mathbb{R}$. We consider a non-degenerate critical point q_0 of index 1, so we can write

$$H(x) = Q(x, x) + R(x),$$

where $R(x) = o(|x|^2)$, so $\lim_{x\to 0} \frac{R(x)}{|x|^2} = 0$.

We first investigate the Hamiltonian $H_Q : x \mapsto Q(x, x)$, and then argue that the results continue to hold qualitatively when R is sufficiently small.

Lemma 12.3.1. *We consider a quadratic Hamiltonian H_Q with a non-degenerate critical point of index* 1*. Fix $c > 0$. Then every level set $H_Q = c$ has a unique simple periodic orbit γ_c (up to reparametrization). This periodic orbit is transversely non-degenerate and its Conley–Zehnder index equals* 2*.*

Proof. By Proposition 8.5.3, we can find symplectic coordinates such that H_Q has the form

$$(\xi_1, \xi_2; \eta_1, \eta_2) \longmapsto -2a\eta_1\eta_2 + b(\xi_1^2 + \xi_2^2)$$

with $a, b > 0$. The symplectic form is given by $d\eta_1 \wedge d\eta_2 + d\xi_1 \wedge d\xi_2$, so the Hamiltonian vector field is given by

$$X_{H_Q} = 2a\partial_{\eta_1} - 2a\partial_{\eta_2} + -b\xi_2\partial_{\xi_1} + b\xi_1\partial_{\xi_2}.$$

This is linear, so we obtain its time-t flow by exponentiating

$$\phi^t_{H_Q}(\eta_1, \eta_2; \xi_1, \xi_2) = \begin{pmatrix} e^{2at} & 0 & 0 & 0 \\ 0 & e^{-2at} & 0 & 0 \\ 0 & 0 & \cos(bt) & -\sin(bt) \\ 0 & 0 & \sin(bt) & \cos(bt) \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \xi_1 \\ \xi_2 \end{pmatrix}.$$

Clearly, if η_1 or η_2 at time $t = 0$ are non-zero, then the solution cannot be periodic: $|\eta_1|$ is strictly increasing when non-zero and $|\eta_2|$ is strictly decreasing when non-zero. This leaves initial conditions with $\eta_1 = \eta_2 = 0$. The solution is clearly periodic in the latter case with period $2\pi/b$. If we fix $s_c = (0, 0; \sqrt{\frac{c}{b}}, 0)$ as initial point, then we find

$$\gamma_c(t) = \sqrt{\frac{c}{b}}(0, 0; \cos(bt), \sin(bt).$$

To see that this orbit is transversely non-degenerate, we note that

$$T_{s_c}\Sigma_c = \mathrm{span}((1, 0, 0, 0), (0, 0, 1, 0), X_H(s_c)).$$

The linearized time-$2\pi/b$ flow acts on the first two vectors as

$$(1, 0, 0, 0) \longmapsto (e^{4\pi a/b}, 0, 0, 0), \quad (0, 1, 0, 0) \longmapsto (0, e^{-4\pi a/b}, 0, 0).$$

Clearly, the restriction of the linearized flow to the span of the first two vectors does not have any eigenvalue equal to 1.

That leaves the Conley–Zehnder index. Rather than explicitly trivializing the contact structure over a disk, we use the following standard trick. First note that the linearized flow extends to a path of symplectic matrices on $(\mathbb{R}^4, \omega_0)$. Namely, with respect to the standard symplectic basis of $\mathbb{R}^4$ we have

$$\psi_{\mathbb{R}^4} : t \longmapsto \begin{pmatrix} e^{2at} & 0 & 0 & 0 \\ 0 & e^{-2at} & 0 & 0 \\ 0 & 0 & \cos(bt) & -\sin(bt) \\ 0 & 0 & \sin(bt) & \cos(bt) \end{pmatrix}.$$

For later use, let us call the path of abstract (not depending on basis) linear symplectic maps the *extended linearized flow* and denote this path by ψ. Note each symplectic matrix $\psi_{\mathbb{R}^4}(t)$ has the direct sum decomposition

$$\psi_{\mathbb{R}^4}(t) = \begin{pmatrix} e^{2at} & 0 \\ 0 & e^{-2at} \end{pmatrix} \oplus \begin{pmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{pmatrix}.$$

We compute the Maslov index, or actually the Robbin–Salamon index, of the former path with the crossing formula, which we have seen in Formula (10.13) for the Lagrangian version and Formula (10.14) for the Conley–Zehnder index. We adapt (10.13) for symplectic paths with a degenerate endpoint. There is only one crossing at $t = 0$, and its signature is zero, so we have

$$\mu\left(\begin{pmatrix} e^{2at} & 0 \\ 0 & e^{-2at} \end{pmatrix}, t = 0, \dots, 2\pi/b \right) = 0.$$

We use the following lemma to compute the Maslov index of the latter path.

Lemma 12.3.2. *Consider a path of symplectic matrices of the form*

$$\Phi : t \longmapsto \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}$$

with t ranging from 0 to T. Then

$$\mu(\Phi) = \lfloor \frac{T}{2\pi} \rfloor + \lceil \frac{T}{2\pi} \rceil.$$

Proof. This formula is most easily proved with Theorem 10.6.1. The endpoint is elliptic, so we only need to compute the rotation number: this equals $\frac{T}{2\pi}$, which then directly gives the result. □

From the crossing formula for the Conley–Zehnder index and Robbin–Salamon index, we see that the following direct sum rule holds for paths of symplectic matrices with the block form of $\psi_{\mathbb{R}^4}$,

$$\begin{aligned}\mu(\psi_{\mathbb{R}^4}) &= \mu\left(\begin{pmatrix} e^{2at} & 0 \\ 0 & e^{-2at}\end{pmatrix}, t=0,\dots,2\pi/b\right) \\ &\quad + \mu\left(\begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t)\end{pmatrix}, t=0,\dots,2\pi/b\right) \\ &= 2.\end{aligned}$$

For the final result we have computed the Maslov index of the second path, the path rotation matrices, with Lemma 12.3.2: this gives 2.

On the other hand, we want to compute the transverse Conley–Zehnder index, of the orbit γ_c. Take a spanning disk $d_c : D^2 \to \Sigma_c$ capping off γ_c. Choose a symplectic trivialization ϵ_ξ of the contact structure ξ along the disk $d_c(D^2)$. Furthermore, we fix the following symplectic trivialization of the symplectic complement of ξ in $T\mathbb{R}^4$ with respect to ω_0,

$$\begin{aligned}\epsilon_{\xi^{\omega_0}} : D^2 \times \mathbb{R}^2 &\longrightarrow d_c^*\xi^{\omega_0} \\ (z; a_1, a_2) &\longmapsto (d_c(z); (a_1 Z \circ d_c(z), a_2 R \circ d_c(z)).\end{aligned}$$

Here Z and R are the Liouville and Reeb vector field, respectively.

We write ψ with respect to the trivialization $\epsilon_\xi \oplus \epsilon_{\xi^{\omega_0}}$, and obtain a path of symplectic matrices

$$\psi_\xi \oplus \psi_{\xi^{\omega_0}}.$$

Since the first Chern class satisfies $c_1(T\mathbb{R}^4) = 0$, the Maslov index of the extended linearized flow ψ does not depend on the choice of the trivialization of $T\mathbb{R}^4$ as we shall see in Lemma 15.2.2. Hence we find

$$\mu(\psi_\xi \oplus \psi_{\xi^{\omega_0}}) = \mu(\psi_{\mathbb{R}^4}) = 2.$$

On the other hand, as mentioned before, the Maslov index of a direct sum of symplectic paths is the sum of the Maslov indices, so

$$\mu(\psi_\xi \oplus \psi_{\xi^{\omega_0}}) = \mu(\psi_\xi) + \mu(\psi_{\xi^{\omega_0}}).$$

We compute the effect of extended linearized flow ψ on vector fields Z and R. We find

$$T\phi^t_{H_Q} Z \circ \gamma_c(0) = Z \circ \gamma_c(t), \quad \text{and} \quad T\phi^t_{H_Q} R \circ \gamma_c(0) = R \circ \gamma_c(t),$$

which means that $\psi_{\xi^{\omega_0}}$ is the constant path. A constant path of symplectic matrices has vanishing Maslov index, so knowing that γ_c is non-degenerate, we conclude that

$$2 = \mu(\psi_\xi \oplus \psi_{\xi^{\omega_0}}) = \mu(\psi_\xi) + \mu(\psi_{\xi^{\omega_0}}) = \mu(\psi_\xi) = \mu_{CZ}(\psi_\xi) = \mu_{CZ}(\gamma_c). \qquad \square$$

We now return to the general case.

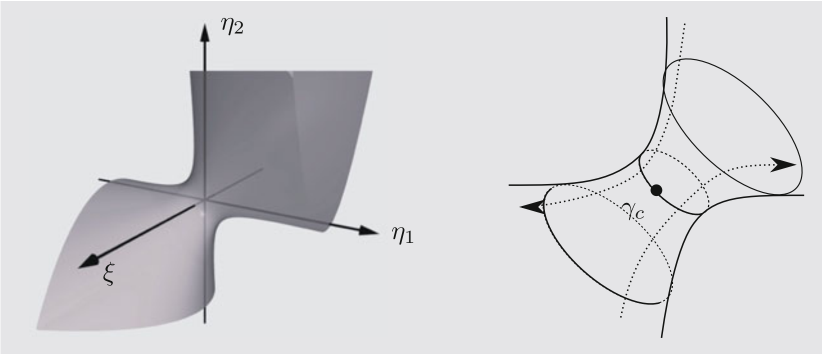


Figure 12.1: Orbits in the tube.

Theorem 12.3.3. *Suppose that (M^4, ω) is a symplectic manifold with Hamiltonian H with a non-degenerate critical point q_0 of index 1 with critical value c_0. Then there is $\epsilon > 0$ such that for all $c \in (c_0, c_0 + \epsilon)$ the energy level set $H^{-1}(c)$ has a simple non-degenerate periodic orbit γ which is transversely non-degenerate and has Conley–Zehnder index equal to 2.*

Proof. In a chart at q_0 we expand the Hamiltonian as $H(x) = c_0 + Q(x,x) + R(x)$, where Q is the Hessian of H at q_0, and $R(x) = o(|x|^2)$. By Lemma 12.3.1, the quadratic part $H_Q(x) = c_0 + Q(x,x)$ has an transversely non-degenerate periodic orbit, say γ_c, with Conley–Zehnder index 2 on each energy level $H_Q^{-1}(c)$ with $c > c_0$.

Since the orbit γ_c is non-degenerate, Theorem 7.7.1 tells us that the orbit will persist under a small perturbation. In other words, for c sufficiently close to c_0, there is an orbit $\tilde{\gamma}_c$ on the energy level $H^{-1}(c)$ corresponding to γ_c.

We claim that the Conley–Zehnder orbit of $\tilde{\gamma}_c$ equals that of γ_c. Here are two ways to see this. Most directly, we can use Formula (10.14) to get a bijection between crossings (including the signature of the crossing form) of the symplectic paths associated with γ_c and $\tilde{\gamma}_c$ provided $\tilde{\gamma}_c$ is a sufficiently small perturbation of γ_c. This holds due to the non-degeneracy. Alternatively, we observe that the perturbed orbit $\tilde{\gamma}_c$ is still hyperbolic and has the same rotation number, again due to non-degeneracy. Theorem 10.6.1 implies the claim. □

Remark 12.3.4. The orbit from the previous theorem is hyperbolic in the sense of Definition 7.7.3 (and the remark after that), and is often called Lyapunov orbit, as we have already seen in Remark 8.5.2.

Chapter 13

Finite Energy Planes

13.1 Holomorphic planes

We assume in this section that (Σ, λ) is a closed, oriented three-dimensional positive contact manifold, i.e., the contact form $\lambda \in \Omega^1(\Sigma)$ satisfies

$$\lambda \wedge d\lambda > 0,$$

or, in other words, the form $\lambda \wedge d\lambda$ is a volume form on Σ inducing the given orientation. The hyperplane distribution

$$\xi = \ker \lambda \subset T\Sigma$$

is referred to as the *contact structure*. The *Reeb vector field* $R \in \Gamma(T\Sigma)$ is implicitly defined by the conditions

$$\lambda(R) = 1, \quad \iota_R d\lambda = 0.$$

The line bundle $\langle R \rangle$ over Σ spanned by the Reeb vector field together with ξ leads to a splitting

$$T\Sigma = \xi \oplus \langle R \rangle$$

of the tangent bundle of Σ. By abuse of notation we extend the contact form λ to a one-form $\lambda \in \Omega^1(\Sigma \times \mathbb{R})$ which at a point $(p, r) \in \Sigma \times \mathbb{R}$ is given by

$$\lambda_{p,r} = e^r \lambda_p.$$

Its differential $\omega = d\lambda$ is a symplectic form for $\Sigma \times \mathbb{R}$. At a point (p, r) it is given by

$$\omega_{p,r} = e^r d\lambda_p + e^r dr \wedge \lambda_p.$$

The non-compact symplectic manifold $(\Sigma \times \mathbb{R}, \omega)$ is called the *symplectization*[4] of the contact manifold (Σ, λ). The vector field ∂_r is a Liouville vector field on

[4]One can argue that one should use $\mathbb{R} \times \Sigma$ for the symplectization instead, since that induces the correct orientation on Σ with standard conventions. We will stick to our choice, though.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_13

the symplectization, indeed, the Lie derivative of ω with respect to ∂_r is given by Cartan's formula by

$$\mathcal{L}_{\partial_r}\omega = d\iota_{\partial_r}\omega + \iota_{\partial_r}d\omega = d\lambda = \omega$$

where we used $\iota_{\partial_r}\omega = \lambda$ and $d\omega = 0$. By abuse of notation we extend the Reeb vector field to $\Gamma(\Sigma \times \mathbb{R})$ by

$$R(p,r) = R(p) \in T_p\Sigma \subset T_p\Sigma \times \mathbb{R} = T_{(p,r)}(\Sigma \times \mathbb{R}), \quad (p,r) \in \Sigma \times \mathbb{R}.$$

Similarly we extend the rank-2 bundle $\xi \subset T\Sigma$ to a rank-2 subbundle $\xi \subset T(\Sigma \times \mathbb{R})$ by

$$\xi_{(p,r)} = \xi_p \subset T_p\Sigma \subset T_p\Sigma \times \mathbb{R} \subset T_{(p,r)}(\Sigma \times \mathbb{R}).$$

We have the splitting

$$T(\Sigma \times \mathbb{R}) = \xi \oplus \langle \partial_r, R \rangle. \tag{13.1}$$

Note that this splitting is symplectic, i.e., both subbundles are symplectic subbundles of $T(\Sigma \times \mathbb{R})$ and they are symplectically orthogonal to each other. Moreover, the symplectic form on ξ is, up to the conformal factor e^r, just the restriction of $d\lambda$ to ξ. Choose $J \in \mathrm{End}(\xi)$ a $d\lambda$-compatible almost complex structure on ξ invariant under the natural $\mathbb{R}$-action

$$\mathbb{R} \times \Sigma \times \mathbb{R} \to \Sigma \times \mathbb{R}, \quad (s,p,r) \mapsto (p, s+r).$$

Again by abuse of notation we extend J to an ω-compatible almost complex structure on $T(\Sigma \times \mathbb{R})$ given for $v \in \xi$ and $a, b \in \mathbb{R}$ by

$$J(v + a\partial_r + bR) = Jv - b\partial_r + aR.$$

Note that the extension is still $\mathbb{R}$-invariant and respects the symplectic splitting (13.1). We refer to such an almost complex structure $J \in \mathrm{End}(T(\Sigma \times \mathbb{R}))$ as an *SFT-like almost complex structure*. Here SFT stands for Symplectic Field theory [77]. We now fix an SFT-like almost complex structure J on $T(\Sigma \times \mathbb{R})$. By a (*parametrized*) *holomorphic plane* we mean a map

$$\widetilde{u}\colon \mathbb{C} \to \Sigma \times \mathbb{R}$$

that is a solution of the following nonlinear Cauchy–Riemann equation

$$\partial_x \widetilde{u} + J(\widetilde{u})\partial_y \widetilde{u} = 0 \tag{13.2}$$

where $z = x + iy$. Recall that the *group of direct similitudes* is the semidirect product

$$\mathscr{S} = \mathbb{C}^* \ltimes \mathbb{C}$$

with multiplication defined as

$$(\rho_1, \tau_1)(\rho_2, \tau_2) = (\rho_1\rho_2, \rho_1\tau_2 + \tau_1).$$

It acts on $\mathbb{C}$ by

$$(\rho,\tau)z = \rho z + \tau, \quad (\rho,\tau) \in \mathbb{C}^* \ltimes \mathbb{C}, \quad z \in \mathbb{C}.$$

Geometrically this amounts to a combination of a rotation, a translation, and a dilation of $\mathbb{C}$. Every biholomorphism of $\mathbb{C}$ to itself is of this form. The group of direct similitudes acts on solutions of (13.2) by reparametrization

$$(\rho,\tau)_*\widetilde{u}(z) = \widetilde{u}(\rho z + \tau).$$

Note that this is actually a right action. We refer to an equivalence class $[\widetilde{u}]$ of a holomorphic plane $\widetilde{u}$ under the action of the group of direct similitudes as an *unparametrized holomorphic plane*.

While the group of direct similitudes acts on the domain of a holomorphic plane, there is an action of the group $\mathbb{R}$ on the target as well. For a holomorphic plane write

$$\widetilde{u} = (u,a), \quad u\colon \mathbb{C} \to \Sigma,\ a\colon \mathbb{C} \to \mathbb{R}.$$

Because the SFT-like almost complex structure J is $\mathbb{R}$-invariant, it follows that for a solution $\widetilde{u}$ of (13.2) and $r \in \mathbb{R}$ the map

$$r_*(u,a) = (u, a+r)\colon \mathbb{C} \to \Sigma \times \mathbb{R}$$

is still a solution of (13.2). Note that the actions of the group $\mathscr{S}$ and $\mathbb{R}$ on solution of (13.2) commute so that we obtain an action of the group $\mathscr{S} \times \mathbb{R}$ on solutions of (13.2). In particular, the group $\mathbb{R}$ still acts on unparametrized holomorphic planes.

13.2 The Hofer energy of a holomorphic plane

In order to describe the *energy* of a holomorphic curve we abbreviate

$$\Gamma := \{\phi \in C^\infty(\mathbb{R},[0,1]) : \phi' \geq 0\}.$$

For $\phi \in \Gamma$ we define $\lambda^\phi \in \Omega^1(\Sigma \times \mathbb{R})$ by

$$\lambda^\phi_{(p,r)} = \phi(r)\lambda_p \quad (p,r) \in \Sigma \times \mathbb{R}.$$

The following notion of energy of a holomorphic plane is due to Hofer [120]

$$E(\widetilde{u}) := \sup_{\phi \in \Gamma} \int_{\mathbb{C}} \widetilde{u}^* d\lambda^\phi. \tag{13.3}$$

Note that the energy is invariant under the action of $\mathscr{S} \times \mathbb{R}$ on solutions of (13.2). For the $\mathbb{R}$-action, this follows from the fact that $\mathbb{R}$ also acts on Γ by

$$r_*\phi(s) = \phi(s-r), \quad s \in \mathbb{R}$$

for $r \in \mathbb{R}$ and $\phi \in \Gamma$. The following lemma tells us that the energy of a holomorphic plane is never negative.

Lemma 13.2.1. *Assume that $\widetilde{u}$ is a solution of the nonlinear Cauchy–Riemann equation* (13.2)*. Then its energy satisfies*

$$E(\widetilde{u}) \in [0, \infty].$$

Moreover, $E(\widetilde{u}) = 0$ if and only if $\widetilde{u}$ is constant.

Proof. Pick $\phi \in \Gamma$. At a point $(p, r) \in \Sigma \times \mathbb{R}$ the exterior derivative of λ^ϕ is given by

$$d\lambda^\phi_{(p,r)} = \phi(r) d\lambda_p + \phi'(r) dr \wedge \lambda_p.$$

Abbreviate by

$$\pi \colon T\Sigma \to \xi$$

the projection along $\langle R \rangle$. By definition of the Reeb vector field, we obtain

$$\partial_x u = \pi \partial_x u + \lambda(\partial_x u) R$$

and therefore

$$\partial_x \widetilde{u} = \pi \partial_x u + \lambda(\partial_x u) R + \partial_x a \partial_r.$$

Using (13.2) we conclude

$$\partial_y \widetilde{u} = J \partial_x \widetilde{u} = J \pi \partial_x u + \partial_x a R - \lambda(\partial_x u) \partial_r.$$

Putting this together we end up with the formula

$$d\lambda^\phi(\partial_x \widetilde{u}, \partial_y \widetilde{u}) = \phi(a) d\lambda(\pi \partial_x u, J \pi \partial_x u) + \phi'(a)\big((\partial_x a)^2 + (\lambda(\partial_x u))^2\big). \tag{13.4}$$

Since the restriction of J to ξ is $d\lambda$-compatible, it follows that $d\lambda(\pi \partial_x u, J \pi \partial_x u) \geq 0$. By definition $\phi(a) \geq 0$ and $\phi'(a) \geq 0$ so that it holds that

$$d\lambda^\phi(\partial_x \widetilde{u}, \partial_y \widetilde{u}) \geq 0.$$

We showed that

$$\int_{\mathbb{C}} \widetilde{u}^* d\lambda^\phi \geq 0, \quad \forall\, \phi \in \Gamma$$

and therefore

$$E(\widetilde{u}) \geq 0.$$

Now assume that $\widetilde{u}$ is not constant. That means that there exists $z \in \mathbb{C}$ such that

$$d\widetilde{u}(z) \neq 0.$$

Since $\widetilde{u}$ satisfies the Cauchy–Riemann equation (13.2) it follows that

$$\partial_x \widetilde{u}(z) \neq 0.$$

It follows from (13.4) that we can choose $\phi \in \Gamma$ such that

$$d\lambda^{\phi}(\partial_x \widetilde{u}, \partial_y \widetilde{u})(z) > 0.$$

Hence the energy satisfies

$$E(\widetilde{u}) > 0.$$

This finishes the proof of the lemma. □

The following definition is due to Hofer [120].

Definition 13.2.2. A holomorphic plane $\widetilde{u}\colon \mathbb{C} \to \Sigma \times \mathbb{R}$ is called a *finite energy plane* if

$$0 < E(\widetilde{u}) < \infty.$$

The motivation of Hofer to study finite energy planes came from its close relation to the Reeb dynamics on the contact manifold (Σ, λ). In the following let $S^1 = \mathbb{R}/\mathbb{Z}$ be the circle and $\mathbb{R}_+ = \{r \in \mathbb{R} : r > 0\}$ the positive real numbers. Recall from Section 7.2 the following definition.

Definition 13.2.3. A (*parametrized*) *periodic orbit of the Reeb vector field R* is a loop $\gamma \in C^\infty(S^1, \Sigma)$ for which there exists $\tau \in \mathbb{R}_+$ such that the pair (γ, τ) is a solution of the problem

$$\partial_t \gamma(t) = \tau R(\gamma(t)), \quad t \in S^1.$$

Because γ is parametrized, the positive number $\tau = \tau(\gamma)$ is uniquely determined by γ and is referred to as the *period* of γ. The following theorem is due to Hofer. To state it we introduce the following notation. If $u\colon \mathbb{C} \to \Sigma$ is a smooth map and $s \in \mathbb{R}$, we abbreviate

$$u^s\colon S^1 \to \Sigma, \quad t \mapsto u(e^{2\pi(s+it)}).$$

Theorem 13.2.4. *Assume that $\widetilde{u} = (u, a)\colon \mathbb{C} \to \Sigma \times \mathbb{R}$ is a finite energy plane. Then there exists a periodic Reeb orbit γ and a sequence $s_k \to \infty$ such that the sequence u^{s_k} converges in the C^∞-topology to γ.*

For a proof of this fundamental theorem also referred to as the main result of finite energy planes we refer to [120, Theorem 31] or [1, Chapter 3]. The original interest in this result came from the fact that it enabled Hofer in [120] to deduce from it the *Weinstein conjecture* for a broad class of three-dimensional contact manifolds. The Weinstein conjecture [241] asks if every closed contact manifold admits a periodic Reeb orbit. The paper by Hofer [120] was one of the important breakthroughs concerning this conjecture. Later on, Taubes [233] proved the Weinstein conjecture in dimension three completely using methods quite different from finite energy planes, namely Seiberg–Witten invariants. In higher dimensions the conjecture is still open in general, and we refer to the paper by Albers and Hofer [10] and the literature cited therein for partial progress in higher dimensions.

The question how far the Weinstein conjecture generalizes to non-compact manifolds is an active topic of research as well. The interested reader might consult the paper by Van den Berg, Pasquotto, and Vandervorst [36] or the paper by Suhr and Zehmisch [228]. An intriguing point is that without the contact condition the conjecture fails for a general Hamiltonian system; for example, the horocycle flow on the unit cotangent bundle of a compact, hyperbolic surface is described by a Hamiltonian consisting of a geodesic flow term plus a magnetic term, yet has no periodic orbits. See the paper by Ginzburg and Gürel [101] and the literature cited therein for more details on related constructions.

13.3 The Omega-limit set of a finite energy plane

Theorem 13.2.4 does not claim that the asymptotic periodic Reeb orbit γ is unique, and in fact, Siefring has shown that this is not the case in general, [221]. The notion describing this is the Omega-limit set $\Omega(u)$ of a finite energy plane $\widetilde{u}$, which we introduce now. Namely, $\Omega(u)$ consists of all periodic Reeb orbits γ for which there exists a sequence s_k going to infinity such that u^{s_k} converges to γ in the C^∞-topology. Note that

$$\Omega(u) \subset C^\infty(S^1, \Sigma)$$

and we topologize it as a subset of the free loop space of Σ. As the notation suggests, the set $\Omega(u)$ only depends on the projection of the finite energy plane $\widetilde{u}$ to Σ. In particular, the Omega-limit set is invariant under the $\mathbb{R}$-action on finite energy planes. Hofer's theorem tells us that the Omega-limit set is never empty.

Lemma 13.3.1. *Assume that $\widetilde{u} = (u, a)$ is a finite energy plane. Then its Omega-limit set $\Omega(u)$ is compact and connected.*

Proof. To prove the lemma we use a statement stronger than the one provided by Theorem 13.2.4. Namely for a given sequence s_k going to infinity there exists a subsequence s_{k_j} and a periodic Reeb orbit γ such that $u^{s_{k_j}}$ converges to γ. However, this improved statement can be shown along the same lines as Theorem 13.2.4, see [1, Theorem 6.4.1] and see also [129, Theorem 3.2]. Armed with this fact we are in position to prove the lemma.

We first show that $\Omega(u)$ is compact. The free loop space $C^\infty(S^1, \Sigma)$ is metrizable. Therefore it suffices to show that $\Omega(u)$ is sequentially compact. Choose a metric d on $C^\infty(S^1, \Sigma)$ which induces the given topology of the free loop space. Let γ_ν for $\nu \in \mathbb{N}$ be a sequence in $\Omega(u)$. Since $\gamma_\nu \in \Omega(u)$ for every $\nu \in \mathbb{N}$, there exists a sequence $\{s_k^\nu\}_{k \in \mathbb{N}}$ going to infinity with the property that

$$\lim_{k \to \infty} u^{s_k^\nu} = \gamma_\nu.$$

Set $k_1 := 1$ and define inductively for $\nu \in \mathbb{N}$

$$k_{\nu+1} := \min \left\{ k : s_k^{\nu+1} \geq s_{k_\nu}^\nu + 1, d\big(u^{s_k^{\nu+1}}, \gamma_{\nu+1}\big) \leq \frac{1}{\nu+1} \right\}.$$

For $\nu \in \mathbb{N}$ define

$$\sigma_\nu := s^\nu_{k_\nu}.$$

It follows by construction that the sequence σ_ν goes to infinity. By the improved version of Hofer's theorem discussed above there exists a subsequence ν_j and a periodic Reeb orbit γ such that

$$\lim_{j\to\infty} u^{\sigma_{\nu_j}} = \gamma.$$

We claim that

$$\lim_{j\to\infty} \gamma_{\nu_j} = \gamma. \tag{13.5}$$

To see that pick $\epsilon > 0$. Choose $j_0 = j_0(\epsilon)$ with the property that

$$\nu_{j_0} \geq \frac{2}{\epsilon}, \quad d(u^{\sigma_{\nu_j}}, \gamma) \leq \frac{\epsilon}{2}, \ \forall\, j \geq j_0.$$

We estimate for every $j \geq j_0$

$$d\big(\gamma_{\nu_j}, \gamma\big) \leq d\big(u^{\sigma_{\nu_j}}, \gamma_{\nu_j}\big) + d\big(u^{\sigma_{\nu_j}}, \gamma\big) \leq \frac{1}{\nu_j} + \frac{\epsilon}{2} \leq \frac{1}{\nu_{j_0}} + \frac{\epsilon}{2} \leq \epsilon.$$

This proves (13.5) and hence $\Omega(u)$ is compact.

It remains to show that $\Omega(u)$ is connected. We assume by contradiction that $\Omega(u)$ is not connected and hence can be written as

$$\Omega(u) = \Omega_1(u) \cup \Omega_2(u)$$

where both $\Omega_1(u)$ and $\Omega_2(u)$ are nonempty, open and closed subsets of $\Omega(u)$ satisfying $\Omega_1(u) \cap \Omega_2(u) = \emptyset$. Since we already know that $\Omega(u)$ is compact the sets $\Omega_1(u)$ and $\Omega_2(u)$ are compact as well and therefore there exist open sets $V_1, V_2 \in C^\infty(S^1, \Sigma)$ with the property that

$$V_1 \cap V_2 = \emptyset, \quad \Omega_1(u) \subset V_1, \quad \Omega_2(u) \subset V_2.$$

Since $\Omega_1(u)$ and $\Omega_2(u)$ are nonempty, there exist $\gamma_1 \in \Omega_1(u)$ and $\gamma_2 \in \Omega_2(u)$. By definition we can find sequences s^1_k and s^2_k going to infinity such that

$$\lim_{k\to\infty} u^{s^1_k} = \gamma_1, \quad \lim_{k\to\infty} u^{s^2_k} = \gamma_2.$$

Set $k_1 = 1$ and define inductively for $\nu \in \mathbb{N}$

$$k_{\nu+1} := \begin{cases} \min\left\{k : s^2_k > s^1_{k_\nu}\right\} & \nu \text{ odd} \\ \min\left\{k : s^1_k > s^2_{k_\nu}\right\} & \nu \text{ even.} \end{cases}$$

Note that the sequence k_ν goes to infinity. For any ν consider the path

$$[s_{k_\nu}, s_{k_{\nu+1}}] \to C^\infty(S^1, \Sigma), \quad s \mapsto u^s.$$

For sufficiently large ν one of the endpoints of this path lies in V_1 while the other one lies in V_2. Therefore there exists $\sigma_\nu \in [s_{k_\nu}, s_{k_\nu+1}]$ with the property that

$$u^{\sigma_\nu} \in C^\infty(S^1, \Sigma) \setminus (V_1 \cap V_2).$$

Observe that the sequence σ_ν goes to infinity since s_{k_ν} goes to infinity. By the improved version of Hofer's theorem there exists a subsequence ν_j and a periodic Reeb orbit such that

$$\lim_{j\to\infty} u^{\sigma_{\nu_j}} = \gamma.$$

By definition of the Omega-limit set we have

$$\gamma \in \Omega(u).$$

On the other hand, V_1 and V_2 were open subsets of the free loop space of Σ and therefore

$$\gamma \in C^\infty(S^1, \Sigma) \setminus (V_1 \cup V_2) \subset C^\infty(S^1, \Sigma) \setminus (\Omega_1(u) \cup \Omega_2(u)) = C^\infty(S^1, \Sigma) \setminus \Omega(u).$$

This contradiction shows that $\Omega(u)$ is connected and the lemma is proved. □

There is a free action of the group S^1 on the set of parametrized periodic Reeb orbits by time-shift. Indeed, if $\gamma \in C^\infty(S^1, \Sigma)$ is a periodic Reeb orbit and $r \in \mathbb{R}/\mathbb{Z}$, the loop $r_*\gamma$ defined as

$$r_*\gamma(t) = \gamma(r + t), \quad t \in S^1$$

is again a periodic Reeb orbit. We refer to an orbit of this action as an unparametrized Reeb orbit, namely

Definition 13.3.2. An *unparametrized Reeb orbit* $[\gamma] = \{r_*\gamma : r \in S^1\}$ is an equivalence class of a parametrized Reeb orbit γ under the equivalence relation given by time-shift.

We say that a periodic Reeb orbit γ is *isolated* if $[\gamma]$ is isolated in the space of unparametrized loops $C^\infty(S^1, \Sigma)/S^1$. Note that a parametrized Reeb orbit can never be isolated in the free loop space $C^\infty(S^1, \Sigma)$ since it always comes in a circle family. If an isolated periodic Reeb orbit γ lies in the Omega-limit set of a finite energy plane $\widetilde{u} = (u, a)$, it follows from Lemma 13.3.1 that

$$\Omega(u) \subset [\gamma].$$

Therefore we abbreviate for an isolated periodic Reeb orbit γ

$$\widehat{\mathcal{M}}(\gamma) := \widehat{\mathcal{M}}([\gamma]) := \big\{\widetilde{u} = (u, a) \text{ finite energy plane, } \Omega(u) \subset [\gamma]\big\} \tag{13.6}$$

the moduli space of finite energy planes asymptotic to the unparametrized periodic orbit $[\gamma]$. Recall that the group of direct similitudes $\mathscr{S} = \mathbb{C}^* \ltimes \mathbb{C}$ acts on finite energy planes by reparametrization. If $(\rho,\tau) \in \mathscr{S}$ with $\rho = |\rho|e^{2\pi i\theta} \in \mathbb{C}^*$ and $\widetilde{u} = (u,a)$ is a finite energy plane, it follows that

$$\Omega((\rho,\tau)_* u) = \theta_* \Omega(u).$$

We conclude that the moduli space $\widehat{\mathcal{M}}(\gamma)$ is invariant under the action of $\mathscr{S}$ and we abbreviate by

$$\mathcal{M}(\gamma) := \widehat{\mathcal{M}}(\gamma)/\mathscr{S}$$

the moduli space of unparametrized finite energy planes asymptotic to $[\gamma]$. Note that since the $\mathbb{R}$-action on finite energy planes given by $r_*(u,a) = (u, a+r)$ commutes with the $\mathscr{S}$-action we still have a $\mathbb{R}$-action on the moduli space of unparametrized finite energy planes.

13.4 Non-degenerate finite energy planes

The situation becomes much nicer if we assume that the periodic Reeb orbit is *non-degenerate*. To explain this notion let us abbreviate by $\phi_R^t \colon \Sigma \to \Sigma$ for $t \in \mathbb{R}$ the flow of the Reeb vector field on Σ defined by

$$\phi_R^0 = \mathrm{id}, \quad \frac{d}{dt}\phi_R^t(x) = R(\phi_R^t(x)),\ x \in \Sigma,\ t \in \mathbb{R}.$$

Note that the contact form λ is invariant under the Reeb flow. Indeed, the Lie derivative of λ with respect to R computes by Cartan's formula to be

$$\mathcal{L}_R \lambda = \iota_R d\lambda + d\iota_R \lambda = 0$$

by the defining equation for the Reeb vector field. It follows that the differential of the Reeb flow

$$d\phi_R^t(x) \colon T_x\Sigma \to T_{\phi_R^t(x)}\Sigma$$

keeps the hyperplane distribution $\xi = \ker \lambda$ invariant so that we can define

$$d^\xi \phi_R^t(x) \colon \xi_x \to \xi_{\phi_R^t(x)}, \quad d^\xi \phi_R^t(x) := d\phi_R^t(x)|_{\xi_x}.$$

Again by the fact that λ and therefore $d\lambda$ are invariant under the Reeb flow we conclude that the map $d^\xi \phi_R^t(x)$ is a linear symplectic map from the symplectic vector space $(\xi_x, d\lambda)$ to the symplectic vector space $(\xi_{\phi_R^t(x)}, d\lambda)$. In particular, if γ is a periodic Reeb orbit of period τ, we obtain a symplectic map

$$d^\xi \phi_R^\tau(\gamma(0)) \colon \xi_{\gamma(0)} \to \xi_{\gamma(0)}.$$

Definition 13.4.1. A periodic Reeb orbit γ of period τ is called *non-degenerate* if

$$\det(d^\xi \phi_R^\tau(\gamma(0)) - \mathrm{id}) \neq 0.$$

Note that if γ is non-degenerate and $[r] \in S^1 = \mathbb{R}/\mathbb{Z}$ the reparametrized periodic orbit $[r]_*\gamma$ is still non-degenerate. Indeed, since ϕ_R^t is a flow we have the relation

$$d^\xi \phi_R^\tau(\gamma(r)) = d^\xi \phi^{\tau r}(\gamma(0)) d^\xi \phi^\tau(\gamma(0)) d^\xi \phi^{\tau r}(\gamma(0))^{-1}.$$

Therefore it makes sense to talk about a non-degenerate unparametrized periodic orbit.

Definition 13.4.2. A finite energy plane $\widetilde{u} = (u, a)$ is called *non-degenerate* if there exists a non-degenerate periodic orbit γ such that $\gamma \in \Omega(u)$.

A non-degenerate periodic orbit γ is isolated and therefore by Lemma 13.3.1 it holds that $\Omega(u) \subset [\gamma]$. However, more can be shown, see [124].

Lemma 13.4.3. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. Then $\Omega(u) = \{\gamma\}$, i.e., the Omega-limit set of a non-degenerate finite energy plane consists of a unique parametrized non-degenerate periodic orbit.*

How a non-degenerate finite energy plane converges to its by the above lemma unique asymptotic orbit has been described quite precisely by Hofer, Wysocki, and Zehnder in [124]. We discuss this in the next section.

Remark 13.4.4. There exists in the literature the more general notion of "non-degenerate puncture", see [133].

13.5 The asymptotic formula

Assume that $\gamma \in C^\infty(S^1, \Sigma)$ is a periodic Reeb orbit of period τ and J is an SFT-like almost complex structure. Denote by $\Gamma^{1,2}(\gamma^*\xi)$ the Hilbert space of $W^{1,2}$-sections in ξ and by $\Gamma^{0,2}(\gamma^*\xi)$ the Hilbert space of L^2-sections in ξ. Consider the bounded linear operator

$$A_\gamma := A_{\gamma,J} \colon \Gamma^{1,2}(\gamma^*\xi) \to \Gamma^{0,2}(\gamma^*\xi)$$

which for $w \in \Gamma^{1,2}(\gamma^*\xi)$ is given by

$$A_\gamma(w)(t) = -J(\gamma(t)) d^\xi \phi_R^{t\tau}(\gamma(0)) \partial_t \Big(d^\xi \phi_R^{-t\tau}(\gamma(t)) w(t) \Big), \quad t \in [0,1].$$

Even though this may not make global sense in the case the orbit is multiply covered, the operator A_γ is basically $-J$ applied to the Lie derivative of a local extension of the section along the flow line to a vector field on some neighborhood of the flow line. Suppose that

$$\mathfrak{T} \colon \gamma^*\xi \to S^1 \times \mathbb{C}$$

is a unitary trivialization, i.e., an orthogonal trivialization, where orthogonality refers to the bundle metric $\omega(\cdot, J\cdot)$ on ξ and the standard inner product on $\mathbb{C}$, which interchanges multiplication by J on $\gamma^*\xi$ with multiplication by i on $\mathbb{C}$. The trivialization $\mathfrak{T}$ gives rise to a Hilbert space isomorphism

$$\Phi_{\mathfrak{T}}\colon \Gamma^{1,2}(\gamma^*\xi) \to W^{1,2}(S^1,\mathbb{C}), \quad w \mapsto \mathfrak{T}w$$

which extends to a Hilbert space isomorphism

$$\Phi_{\mathfrak{T}}\colon \Gamma^{0,2}(\gamma^*\xi) \to L^2(S^1,\mathbb{C})$$

by the same formula. Hence we obtain an operator

$$A_\gamma^{\mathfrak{T}} := \Phi_{\mathfrak{T}} A_\gamma \Phi_{\mathfrak{T}}^{-1}\colon W^{1,2}(S^1,\mathbb{C}) \to L^2(S^1,\mathbb{C}). \tag{13.7}$$

To describe $A_\gamma^{\mathfrak{T}}$ we write J_0 for the standard complex structure on $\mathbb{C}$ given by multiplication with i. For $t \in [0,1]$ we further abbreviate

$$\Psi(t) := \mathfrak{T}_{\gamma(t)} d^\xi \phi_R^{t\tau}(\gamma(0)) \mathfrak{T}_{\gamma(0)}^{-1} \in \mathrm{Sp}(1)$$

as well as

$$S(t) = -J_0 \partial_t \Psi(t)\Psi(t)^{-1} \in \mathrm{Sym}(2).$$

Assume $v \in W^{1,2}(S^1,\mathbb{C})$ and $t \in [0,1]$. We compute

$$\begin{aligned}
(A_\gamma^{\mathfrak{T}})v(t) &= -\mathfrak{T}_{\gamma(t)} J(\gamma(t)) d^\xi \phi_R^{t\tau}(\gamma(0)) \partial_t \Big(d^\xi \phi_R^{-t\tau}(\gamma(t)) \mathfrak{T}_{\gamma(t)}^{-1} v(t) \Big) \\
&= -J_0 \mathfrak{T}_{\gamma(t)} d^\xi \phi_R^{t\tau}(\gamma(0)) \partial_t \Big(d^\xi \phi_R^{-t\tau}(\gamma(t)) \mathfrak{T}_{\gamma(t)}^{-1} v(t) \Big) \\
&= -J_0 \mathfrak{T}_{\gamma(t)} d^\xi \phi_R^{t\tau}(\gamma(0)) \mathfrak{T}_{\gamma(0)}^{-1} \partial_t \Big(\mathfrak{T}_{\gamma(0)} d^\xi \phi_R^{-t\tau}(\gamma(t)) \mathfrak{T}_{\gamma(t)}^{-1} v(t) \Big) \\
&= -J_0 \Psi(t) \partial_t \Big(\Psi(t)^{-1} v(t) \Big) \\
&= -J_0 \partial_t v(t) + J_0 \partial_t \Psi(t) \Psi(t)^{-1} v(t) \\
&= -J_0 \partial_t v(t) - S(t) v(t).
\end{aligned}$$

That means that

$$A_\gamma^{\mathfrak{T}} = A_S,$$

where A_S is the operator defined in (11.4). Since the operator A_γ is conjugated to the operator $A_\gamma^{\mathfrak{T}}$, it has the same spectral properties as A_S. In particular, its spectrum is discrete and consists of real eigenvalues of finite multiplicity.

The following notion is due to Siefring [219].

Definition 13.5.1. Assume that $\widetilde{u}$ is a non-degenerate finite energy plane with asymptotic orbit γ of period τ and $U\colon [R,\infty) \times S^1 \to \gamma^*\xi$ is a smooth map such that $U(s,t) \in \xi_{\gamma(t)}$ for all $(s,t) \in [R,\infty) \times S^1$. The map U is called an *asymptotic representative* if there exists a proper embedding $\phi\colon [R,\infty) \times S^1 \to \mathbb{R} \times S^1$ asymptotic to the identity such that

$$\widetilde{u}(e^{\phi(s,t)}) = (\exp_{\gamma(t)} U(s,t), \tau s)$$

where exp is the exponential map of the restriction of the metric $\omega(\cdot, J\cdot)$ to $\Sigma = \Sigma \times \{0\} \subset \Sigma \times \mathbb{R}$.

The following theorem follows from results of Mora [184] based on previous work by Hofer, Wysocki, and Zehnder [124].

Theorem 13.5.2. *Assume that the map $\widetilde{u}$ is a non-degenerate finite energy plane with asymptotic orbit γ. Then $\widetilde{u}$ admits an asymptotic representative. Moreover, there exist a negative eigenvalue η of A_γ and an eigenvector ζ of A_γ to the eigenvalue η such that the asymptotic representative can be written as*

$$U(s,t) = e^{\eta s}\big(\zeta(t) + \kappa(s,t)\big) \tag{13.8}$$

where κ decays exponentially with all derivatives in the sense that there exist for one and hence every metric constants $M_{i,j}$ for $0 \leq i, j < \infty$ and $d > 0$ such that

$$|\nabla_s^i \nabla_t^j \kappa(s,t)| \leq M_{i,j} e^{-ds}.$$

In view of the requirement that the coordinate change of an asymptotic representative is asymptotic to the identity, an asymptotic representative is unique up to restriction of the domain of definition. In particular, the eigenvalue η and the eigenvector ζ are uniquely determined by the finite energy plane $\widetilde{u}$. We denote the eigenvalue by

$$\eta_{\widetilde{u}} \in \mathfrak{S}(A_\gamma) \cap (-\infty, 0)$$

and refer to it as the *asymptotic eigenvalue* and similarly we denote the eigenvector by

$$\zeta_{\widetilde{u}} \in \Gamma(\gamma^* \xi)$$

and refer to it as the *asymptotic eigenvector*.

Recall that if $\widetilde{u} = (u, a)$ is a finite energy plane, then the group $\mathbb{R}$ acts on it by $r_* \widetilde{u} = (u, a + r)$. For later reference it will be useful to know how the asymptotic eigenvalue and the asymptotic eigenvector transform under this action. To compute this, let U be an asymptotic representative of $\widetilde{u}$ for a proper embedding $\phi\colon [R, \infty) \times S^1 \to \mathbb{R} \times S^1$ asymptotic to the identity. As usual τ stands for the period of the asymptotic Reeb orbit. For $r \in \mathbb{R}$ define

$$U_r\colon [R + \tfrac{r}{\tau}, \infty) \times S^1 \to \gamma^* \xi, \quad (s,t) \mapsto U(s - \tfrac{r}{\tau}, t)$$

and

$$\phi_r\colon [R + \tfrac{r}{\tau}, \infty) \times S^1 \to \mathbb{R} \times S^1, \quad (s,t) \mapsto \phi(s - \tfrac{r}{\tau}, t).$$

We compute

$$\begin{aligned} r_* \widetilde{u}(e^{\phi_r(s,t)}) &= r_* \widetilde{u}(e^{\phi(s - \frac{r}{\tau}, t)}) \\ &= (\exp_{\gamma(t)} U(s - \tfrac{r}{\tau}, t), \tau(s - \tfrac{r}{\tau}) + r) \\ &= (\exp_{\gamma(t)} U_r(s,t), \tau s). \end{aligned}$$

Using (13.8) we compute for U_r

$$\begin{aligned} U_r(s,t) &= U(s-\tfrac{r}{\tau},t) \\ &= e^{\eta(s-\frac{r}{\tau})}\Big(\zeta(t)+\kappa(s-\tfrac{r}{\tau},t)\Big) \\ &= e^{\eta s}\Big(e^{-\frac{\eta r}{\tau}}\zeta(t)+e^{-\frac{\eta r}{\tau}}\kappa(s-\tfrac{r}{\tau},t)\Big). \end{aligned}$$

That means that the asymptotic eigenvalue is unchanged under the $\mathbb{R}$-action while the asymptotic eigenvector gets scaled by a factor $e^{-\frac{\eta r}{\tau}}$. We summarize this computation in the following lemma.

Lemma 13.5.3. *Assume that $\widetilde{u}=(u,a)$ is a non-degenerate finite energy plane with asymptotic orbit γ of period τ. The asymptotic eigenvalue only depends on the projection u, i.e.,*

$$\eta_u := \eta_{\widetilde{u}}$$

while the asymptotic eigenvector transforms under the $\mathbb{R}$-action on $\widetilde{u}$ as

$$\zeta_{r_*\widetilde{u}} = e^{-\frac{\eta_u r}{\tau}}\zeta_{\widetilde{u}}.$$

An important corollary of Theorem 13.5.2 is the following result.

Corollary 13.5.4. *If γ is a non-degenerate Reeb orbit, then the action of $\mathbb{R}$ on the moduli space $\widehat{\mathcal{M}}(\gamma)$ of finite energy planes asymptotic to γ is a free action.*

Proof. Pick a finite energy plane $\widetilde{u}=(u,a)\in\widehat{\mathcal{M}}(\gamma)$. In view of the fact that $\widetilde{u}$ admits an asymptotic representative, it follows that the infimum of the function $a\colon \mathbb{C}\to\mathbb{R}$ is attained and we set

$$\underline{a} := \min\{a(z) : z\in\mathbb{C}\}\in\mathbb{R}.$$

Now suppose that $r\in\mathbb{R}$ satisfies $r_*\widetilde{u}=\widetilde{u}$. Since $r_*\widetilde{u}=(u,a+r)$, we obtain

$$\underline{a}+r=\underline{a}$$

implying that $r=0$. This proves that the $\mathbb{R}$-action on $\widehat{\mathcal{M}}(\gamma)$ is free. $\square$

Remark 13.5.5. In fact, it can be shown that the $\mathbb{R}$-component of a finite energy plane always goes to infinity as $s\to\infty$, even without the non-degeneracy assumption. Therefore, the key fact used in the proof of Corollary 13.5.4 – that the $\mathbb{R}$-component of a finite energy plane always has a minimum – is true in general and does not depend on Theorem 13.5.2 or the assumption of non-degeneracy.

13.6 The index inequality and fast finite energy planes

We first define the Conley–Zehnder index of a non-degenerate finite energy plane $\widetilde{u} = (u, a)$ with asymptotic orbit γ of period τ. Consider the symplectic bundle $u^*\xi \to \mathbb{C}$. Since $\mathbb{C}$ is contractible there exists a symplectic trivialization

$$\mathfrak{T} \colon u^*\xi \to \mathbb{C} \times \mathbb{C}.$$

In view of the asymptotic behavior of $\widetilde{u}$ explained in Theorem 13.5.2 we can arrange the trivialization such that it extends asymptotically to a symplectic trivialization

$$\mathfrak{T} \colon \gamma^*\xi \to S^1 \times \mathbb{C}.$$

Recall that $d^\xi \phi_R^t(\gamma(0)) \colon \xi_{\gamma(0)} \to \xi_{\gamma(t)}$ denotes the restriction of the differential of the Reeb flow to the hyperplane distribution, which turns out to be a linear symplectic map. Hence we obtain a smooth path $\Psi \colon [0,1] \to \mathrm{Sp}(1)$ of symplectic maps from $\mathbb{C}$ to itself by

$$\Psi(t) = \mathfrak{T}_{\gamma(t\tau)} d^\xi \phi_R^{t\tau}(\gamma(0)) \mathfrak{T}_{\gamma(0)}^{-1}.$$

Note that $\Psi(0) = \mathrm{id}$. Moreover, due to the assumption that the asymptotic Reeb orbit γ is non-degenerate, the path of symplectic maps Ψ is non-degenerate in the sense that

$$\ker(\Psi(1) - \mathrm{id}) = \{0\}.$$

We define the Conley–Zehnder index of $\widetilde{u}$ as

$$\mu_{CZ}(\widetilde{u}) = \mu_{CZ}(\Psi). \tag{13.9}$$

Note that since every two symplectic trivializations over the contractible base $\mathbb{C}$ are homotopic, it follows that the Conley–Zehnder index does not depend on the choice of the trivialization $\mathfrak{T}$. Moreover, it depends only on the projection u so that we can also write

$$\mu_{CZ}(u) := \mu_{CZ}(\widetilde{u}).$$

The following index inequality is due to Hofer, Wysocki, and Zehnder [122]

Theorem 13.6.1. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. Then its Conley–Zehnder index satisfies the inequality*

$$\mu_{CZ}(u) \geq 2.$$

Before starting with the proof of this theorem, let us explain how this theorem can be interpreted as an *automatic transversality result*. Denote by γ the asymptotic Reeb orbit of $\widetilde{u}$. Then we can think of $\widetilde{u}$ as an element of the moduli space $\widehat{\mathcal{M}}(\gamma)$ as explained in (13.6). The unparametrized finite energy plane $[\widetilde{u}]$ is then an element of the moduli space $\mathcal{M}(\gamma) = \widehat{\mathcal{M}}(\gamma)/\mathscr{S}$ where $\mathscr{S} = \mathbb{C}^* \ltimes \mathbb{C}$ is the group of direct similitudes which acts on $\widehat{\mathcal{M}}(\gamma)$ by reparametrizations. We will see later in (15.27) that the virtual dimension of the moduli space $\mathcal{M}(\gamma)$ at $[\widetilde{u}]$ is given by

$$\mathrm{virdim}_{[\widetilde{u}]}\mathcal{M}(\gamma) = \mu_{CZ}(u) - 1.$$

Here the virtual dimension of the moduli space $\mathcal{M}(\gamma)$ is given by the Fredholm index of a Fredholm operator L which linearizes the holomorphic curve equation in the normal direction of $\widetilde{u}$. If the Fredholm operator L is surjective locally around $[\widetilde{u}]$, the moduli space $\mathcal{M}(\gamma)$ is a manifold and the tangent space of $\mathcal{M}(\gamma)$ at $[\widetilde{u}]$ equals

$$T_{[\widetilde{u}]}\mathcal{M}(\gamma) = \ker L.$$

Hence if L is surjective, we have

$$\mathrm{virdim}_{[\widetilde{u}]}\mathcal{M}(\gamma) = \mathrm{ind}L = \dim\,\ker L = \dim T_{[\widetilde{u}]}\mathcal{M}(\gamma).$$

Recall that the group $\mathbb{R}$ acts on finite energy planes by $r_*(u, a) = (u, a + r)$. Since this action commutes with the reparametrization action of the group $\mathscr{S}$, the group $\mathbb{R}$ still acts on the moduli space $\mathcal{M}(\gamma)$. Moreover, by Corollary 13.5.4 this action is free. Therefore, still assuming that L is surjective, we get the inequality

$$\dim T_{[\widetilde{u}]}\mathcal{M}(\gamma) \geq 1.$$

Combining these facts we end up with the inequality $\mu_{CZ}(u) \geq 2$ claimed in Theorem 13.6.1. However, we point out that this reasoning only works under the assumption that L is surjective. Geometrically the linearization operator L can be thought of as follows. One interprets the moduli space $\mathcal{M}(\gamma)$ as the zero set of a section

$$s\colon \mathcal{B} \to \mathcal{E}, \quad \mathcal{M}(\gamma) = s^{-1}(0).$$

of a space $\mathcal{B}$ into a bundle $\mathcal{E}$ over $\mathcal{B}$. The Fredholm operator L then arises as the vertical differential of the section s at $[\widetilde{u}]$ and the question if L is surjective can be rephrased geometrically as the question if the section s is transverse to the zero section at $[\widetilde{u}]$. That explains why one refers to Theorem 13.6.1 as an automatic transversality result.

We mention that many transversality results for moduli spaces are so-called generic transversality results which hold for a generic choice of data. In our set-up the data is the SFT-like almost complex structure J. It is therefore highly remarkable that Theorem 13.6.1 holds for any SFT-like almost complex structure J and not just for a generic choice of it.

We now start with the preparations for the proof of Theorem 13.6.1. We assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. We choose a trivialization

$$\mathfrak{T}\colon u^*\xi \to \mathbb{C} \times \mathbb{C}.$$

We further denote by

$$\pi\colon T\Sigma = \xi \oplus \langle R\rangle \to \xi$$

the projection along the Reeb vector field R. A major ingredient in the proof of Theorem 13.6.1 is the map

$$\mathfrak{I}\pi\partial_x u\colon \mathbb{C} \to \mathbb{C}. \tag{13.10}$$

Here we denote by $z = x+iy$ the coordinates on $\mathbb{C}$. This map is smooth interpreted as a real map from $\mathbb{C} = \mathbb{R}^2$ to itself. Moreover, due to the fact that u is holomorphic the map above is "almost holomorphic" in a sense to be described more precisely below.

We first recall some facts about winding for a general smooth map $f\colon \mathbb{C} \to \mathbb{C}$. We denote the regular set of f by

$$\mathcal{R}_f := \{z \in \mathbb{C} : f(z) \neq 0\}.$$

For a continuous loop $\gamma\colon S^1 \to \mathcal{R}_f$ the map $t \mapsto \frac{f(\gamma(t))}{||f(\gamma(t))||}$ is a continuous map from the circle S^1 to itself. Hence we can consider its degree

$$w_\gamma(f) := \deg\Big(t \mapsto \tfrac{f(\gamma(t))}{||f(\gamma(t))||}\Big) \in \mathbb{Z}. \tag{13.11}$$

We refer to $w_\gamma(f)$ as the *winding number* of f along the loop γ. It has the following properties.

Homotopy invariance: If $\gamma\colon S^1 \times [0,1] \to \mathcal{R}_f$ is a continuous map, then $\gamma_0 = \gamma(\cdot, 0)$ and $\gamma_1 = \gamma(\cdot, 1)$ are two homotopic loops in $\mathcal{R}_f$ and its winding numbers are unchanged

$$w_{\gamma_0}(f) = w_{\gamma_1}(f).$$

Concatenation: Suppose that $\gamma_1\colon S^1 \to \mathcal{R}_f$ and $\gamma_2\colon S^1 \to \mathcal{R}_f$ are two continuous maps satisfying $\gamma_1(0) = \gamma_2(0)$. Denote by $\gamma_1\#\gamma_2$ its concatenation. The winding number is additive under concatenation

$$w_{\gamma_1\#\gamma_2}(f) = w_{\gamma_1}(f) + w_{\gamma_2}(f).$$

These two properties have the following consequences. Assume that the singular set

$$\mathcal{S}_f := \{z \in \mathbb{C} : f(z) = 0\} = \mathbb{C} \setminus \mathcal{R}_f$$

is discrete. Pick $z_0 \in \mathcal{S}_f$. Because $\mathcal{S}_f$ is discrete, there exists $\epsilon > 0$ such that the intersection of $\mathcal{S}_f$ with $D_\epsilon(z_0) = \{z \in \mathbb{C} : ||z - z_0|| \leq \epsilon\}$, the closed ϵ-ball around z_0, satisfies

$$\mathcal{S}_f \cap D_\epsilon(z_0) = \{z_0\}.$$

Consider

$$\gamma^\epsilon_{z_0}\colon S^1 \to \mathcal{R}_f, \quad t \mapsto z_0 + \epsilon e^{2\pi i t}.$$

Define the *local winding number* of f at the singularity z_0 by

$$w_{z_0}(f) := w_{\gamma^\epsilon_{z_0}}(f).$$

Because of homotopy invariance of the winding number the local winding number is well defined, independent of the choice of ϵ.

Suppose that $R > 0$ and that $f(Re^{2\pi it}) \neq 0$ for every $t \in S^1$, i.e., we get a loop in the regular set of f

$$\gamma^R \colon S^1 \to \mathcal{R}_f, \quad t \mapsto Re^{2\pi it}.$$

We set

$$w_R(f) := w_{\gamma^R}(f). \tag{13.12}$$

In view of the homotopy invariance and the concatenation property of the winding number we can express this winding number as the sum of local winding numbers

$$w_R(f) = \sum_{z \in D_R(0) \cap \mathcal{S}_f} w_z(f). \tag{13.13}$$

Suppose now that $f \colon \mathbb{C} \to \mathbb{C}$ is *holomorphic* and does not vanish identically. We claim that if $f(z_0) = 0$, then the local winding number satisfies

$$w_{z_0}(f) \geq 1. \tag{13.14}$$

To prove this inequality we can assume without loss of generality that $z_0 = 0$. Since f is holomorphic, it is given by its Taylor series

$$f(z) = \sum_{n=1}^{\infty} a_n z^n.$$

Abbreviate by

$$\ell := \min\{n : a_n \neq 0\}$$

the order of vanishing of f at zero. Since f does not vanish identically, this number ℓ is finite. We can write f now as

$$f(z) = \sum_{n=\ell}^{\infty} a_n z_n.$$

Choose $\epsilon > 0$ such that

$$\sum_{n=\ell+1}^{\infty} |a_n| \epsilon^{n-\ell} < |a_\ell|.$$

Abbreviate

$$g(z) := \sum_{n=\ell+1}^{\infty} a_n z^n$$

so that we obtain

$$f(z) = a_\ell z^\ell + g(z).$$

Consider the map from S^1 to S^1

$$t \mapsto \frac{a_\ell \epsilon^\ell e^{2\pi i t\ell} + g(\epsilon e^{2\pi i t\ell})}{||a_\ell \epsilon^\ell e^{2\pi i t\ell} + g(\epsilon e^{2\pi i t\ell})||}.$$

This map is homotopic to the map from S^1 to S^1 given by

$$t \mapsto \frac{a_\ell \epsilon^\ell e^{2\pi i t\ell}}{||a_\ell \epsilon^\ell e^{2\pi i t\ell}||} = e^{2\pi i t\ell + i\theta}$$

where

$$a_\ell = |a_\ell| e^{i\theta}.$$

This implies that

$$w_0(f) = \ell \geq 1.$$

In other words, the local winding number of a holomorphic function at a zero is given by the order of vanishing of the function at this point. In combination with (13.13) we have established the following lemma.

Lemma 13.6.2. *Assume that $f\colon \mathbb{C} \to \mathbb{C}$ is holomorphic and $R > 0$ such that $f(Re^{2\pi i t}) \neq 0$ for every $t \in S^1$. Then $w_R(f) \geq 0$ and $w_R(f) = 0$ if and only if $f(z) \neq 0$ for every $z \in D_R(0)$.*

We cannot apply Lemma 13.6.2 to the map $\mathfrak{I}\pi\partial_x u\colon \mathbb{C} \to \mathbb{C}$ from (13.10) because this map is not actually holomorphic despite the fact that $\widetilde{u}$ was a holomorphic plane. However, $\mathfrak{I}\pi\partial_x u$ is close enough to be holomorphic that the reasoning which established Lemma 13.6.2 still applies. This is made precise by Carleman's similarity principle. We first examine how far the map $\mathfrak{I}\pi\partial_x u$ deviates from being holomorphic.

From the Darboux theorem, see Theorem 2.5.3, we know that the contact form λ looks locally like $dq_1 + q_2 dq_3$, the standard contact form on $\mathbb{R}^3$. Since the question if the local winding numbers of the map $\mathfrak{I}\pi\partial_x u$ are positive is a local problem, we can assume in view of Darboux's theorem without loss of generality that

$$\lambda = dq_1 + q_2 dq_3, \quad d\lambda = dq_2 \wedge dq_3.$$

In these coordinates the Reeb vector field is given by

$$R = \partial_{q_1}.$$

Moreover, a basis of the hyperplane distribution $\xi = \ker \lambda$ is given by the vectors

$$e_1 = \partial_{q_2}, \quad e_2 = -q_2 \partial_{q_1} + \partial_{q_3}.$$

Note that

$$d\lambda(e_1, e_2) = 1$$

so that the basis $\{e_1, e_2\}$ is a symplectic basis of ξ. If we write

$$u(x,y) = \big(q_1(x,y), q_2(x,y), q_3(x,y)\big)$$

we have

$$\partial_x u(x,y) = \big(\partial_x q_1(x,y), \partial_x q_2(x,y), \partial_x q_3(x,y)\big)$$

and therefore

$$\begin{aligned}\pi_{u(x,y)}\partial_x u(x,y) &= \big(-q_2(x,y)\partial_x q_3(x,y), \partial_x q_2(x,y), \partial_x q_3(x,y)\big)\\ &= \partial_x q_2(x,y) e_1\big(u(x,y)\big) + \partial_x q_3(x,y) e_2\big(u(x,y)\big).\end{aligned}$$

Because all trivializations on a ball are homotopic we can in view of the homotopy invariance of the winding number choose an arbitrary local trivialization to compute the local winding number. We choose as local trivialization the symplectic trivialization

$$\mathfrak{T}_u \colon \xi_u \to \mathbb{C}, \quad (xe_1 + ye_2) \mapsto x + iy.$$

In this trivialization we obtain

$$\mathfrak{T}\pi\partial_x u = \partial_x q_2 + i\partial_x q_3.$$

Because u is holomorphic, the projection to the hyperplane distribution satisfies the equation

$$\pi\partial_x u + J(u)\pi\partial_y u = 0. \tag{13.15}$$

Similar to the above computation for $\pi_u\partial_x u$, we get

$$\pi_u\partial_y u = \partial_y q_2 e_1(u) + \partial_y q_3 e_2(u).$$

We denote by $J(x,y)$ the 2×2-matrix representing $J(u(x,y))$ with respect to the basis $\{e_1(u(x,y)), e_2(u(x,y))\}$. In particular, we have

$$J(x,y)^2 = -\mathrm{id}.$$

With this notation (13.15) translates into

$$\begin{pmatrix} \partial_x q_2 \\ \partial_x q_3 \end{pmatrix} + J(x,y)\begin{pmatrix} \partial_y q_2 \\ \partial_y q_3 \end{pmatrix} = 0.$$

This implies

$$\begin{aligned}\partial_x(\mathfrak{T}\pi\partial_x u) &= \begin{pmatrix} \partial_x^2 q_2 \\ \partial_x^2 q_3 \end{pmatrix}\\ &= -\partial_x J\begin{pmatrix} \partial_y q_2 \\ \partial_y q_3 \end{pmatrix} - J\begin{pmatrix} \partial_x\partial_y q_2 \\ \partial_x\partial_y q_3 \end{pmatrix}\\ &= -(\partial_x J)J\begin{pmatrix} \partial_x q_2 \\ \partial_x q_3 \end{pmatrix} - J\begin{pmatrix} \partial_y\partial_x q_2 \\ \partial_y\partial_x q_3 \end{pmatrix}.\end{aligned}$$

Abbreviating

$$A = -(\partial_x J)J$$

we can write this as

$$\partial_x(\mathfrak{T}\pi\partial_x u) = -A(\mathfrak{T}\pi\partial_x u) - J\partial_y(\mathfrak{T}\pi\partial_x u)$$

or equivalently

$$\partial_x(\mathfrak{T}\pi\partial_x u) + J\partial_y(\mathfrak{T}\pi\partial_x u) + A(\mathfrak{T}\pi\partial_x u) = 0. \tag{13.16}$$

We recall Carleman's similarity principle from [86].

Lemma 13.6.3 (Carleman's similarity principle)**.** *Assume that $f\colon B_\epsilon = \{z \in \mathbb{R}^2 : |z| < \epsilon\} \to \mathbb{R}^2$ is a smooth map and $J, A \in C^\infty(B_\epsilon, M_2(\mathbb{R}))$ are smooth families of 2×2-matrices such that $J(z)^2 = -\mathrm{id}$ for every $z \in B_\epsilon$, i.e., $J(z)$ is a complex structure. Suppose that*

$$\partial_x f + J\partial_y f + Af = 0, \quad f(0) = 0.$$

Then there exists $\delta \in (0, \epsilon)$, $\Phi \in C^0(B_\delta, \mathrm{GL}(\mathbb{R}^2))$, and $\sigma\colon B_\delta \to \mathbb{C}$ holomorphic such that

$$f(z) = \Phi(z)\sigma(z), \quad \sigma(0) = 0, \quad J(z)\Phi(z) = \Phi(z)i.$$

In view of (13.16) and Carleman's similarity principle, all local winding numbers of $\mathfrak{T}\pi\partial_x u$ are positive. Therefore the same reasoning which we used for Lemma 13.6.2 establishes the following proposition.

Proposition 13.6.4. *Assume that $R > 0$ such that $\pi\partial_x u(Re^{2\pi it}) \neq 0$ for every $t \in S^1$. Then $w_R(\mathfrak{T}\pi\partial_x u) \geq 0$ and $w_R(\mathfrak{T}\pi\partial_x u) = 0$ if and only if $\pi\partial_x u(z) \neq 0$ for every $z \in D_R(0)$.*

Note that since an eigenvector ζ of the asymptotic operator A_γ is a solution to a linear ODE, it follows that $\zeta(t) \neq 0$ for every $t \in S^1$. Therefore, in view of the asymptotic behavior of a non-degenerate finite energy plane explained in Theorem 13.5.2 there exists $R_0 > 0$ such that for all $R \geq R_0$ it holds that $\pi\partial_x u(Re^{2\pi it}) \neq 0$ for every $t \in S^1$. Moreover, again in view of the asymptotic behavior, it holds that

$$w_R(\mathfrak{T}\pi\partial_r u) = w(\eta_u) \tag{13.17}$$

where η_u is the asymptotic eigenvalue of the non-degenerate finite energy plane $\widetilde{u}$. Because η_u is negative, this fact has interesting consequences as we will see soon. However, let us first discuss how $w_R(\mathfrak{T}\pi\partial_x u)$ and $w_R(\mathfrak{T}\pi\partial_r u)$ are related. To see that note that

$$\mathfrak{T}\pi\partial_x u = (\mathfrak{T}\pi du)\partial_x, \quad \mathfrak{T}\pi\partial_r u = (\mathfrak{T}\pi du)\partial_r$$

and for each $z \in \mathbb{C}$ the map $\mathfrak{T}\pi du(z)$ is a linear map from $\mathbb{C} = \mathbb{R}^2$ to itself. In general, if $A \in C^\infty(S^1, GL(\mathbb{R}^2))$ and $v_1, v_2 \in C^\infty(S^1, \mathbb{R}^2 \setminus \{0\})$ the formula

$$\deg\left(t \mapsto \frac{A(t)v_2(t)}{|A(t)v_2(t)|}\right) = \deg\left(t \mapsto \frac{A(t)v_1(t)}{|A(t)v_1(t)|}\right) + \mathrm{wind}(v_1, v_2) \tag{13.18}$$

holds true, where $\operatorname{wind}(v_1, v_2)$ is the winding of v_2 around v_1. Because

$$\operatorname{wind}(\partial_x, \partial_r) = 1$$

we obtain the relation

$$w_R(\mathfrak{T}\pi\partial_r u) = w_R(\mathfrak{T}\pi\partial_x u) + \operatorname{wind}(\partial_x, \partial_r) = w_R(\mathfrak{T}\pi\partial_x u) + 1. \tag{13.19}$$

Combining Proposition 13.6.4 with (13.17) and (13.19) we obtain

Theorem 13.6.5. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. Then the winding number of its asymptotic eigenvalue meets the inequality*

$$w(\eta_u) \geq 1.$$

Moreover, $w(\eta_u) = 1$ if and only if $\pi\partial_x u(z) \neq 0$ for every $z \in \mathbb{C}$.

Theorem 13.6.1 is a straightforward consequence of Theorem 13.6.5.

Proof of Theorem 13.6.1. Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. It follows from Theorem 13.6.5 that the winding number of its asymptotic eigenvalue satisfies $w(\eta_u) \geq 1$. Because η_u is negative, it follows from the definition of α in (11.13) that

$$\alpha \geq 1.$$

Because the parity (11.14) satisfies $p \in \{0, 1\}$, we obtain from Theorem 11.3.3 that

$$\mu_{CZ}(u) = 2\alpha + p \geq 2\alpha \geq 2.$$

This finishes the proof of Theorem 13.6.1. □

Assume that $\widetilde{u} = (u, a)$ is a finite energy plane and $\pi\partial_x u(z) \neq 0$ for every $z \in \mathbb{C}$. Because $\widetilde{u}$ is holomorphic, it holds that $\pi\partial_x u + J(u)\pi\partial_y u = 0$. Hence, because $\pi\partial_x u(z) \neq 0$, we have that $\{\pi\partial_x u(z), \pi\partial_y u(z)\}$ are linearly independent vectors in $\xi_{u(z)}$. This implies that

$$\pi du(z) \colon T_z\mathbb{C} = \mathbb{C} \to \xi_{u(z)}$$

is bijective. In particular, $du(z) \colon T_z\mathbb{C} \to T_{u(z)}\Sigma$ is injective and therefore u is an immersion. Moreover, since $T\Sigma = \xi \oplus \langle R \rangle$, the Reeb vector field is transverse to the image of u. On the other hand, if u is an immersion transverse to the Reeb vector field, we must have $\pi\partial_x u(z) \neq 0$ for every $z \in \mathbb{C}$. Therefore we obtain from Theorem 13.6.5 the following corollary.

Corollary 13.6.6. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. Then the winding number of its asymptotic eigenvalues satisfies $w(\eta_u) = 1$ if and only if u is an immersion transverse to the Reeb vector field.*

The following definition is due to Hryniewicz [133].

Definition 13.6.7. A non-degenerate finite energy plane $\widetilde{u} = (u, a)$ is called *fast* if and only if u is an immersion transverse to the Reeb vector field.

The reason for this terminology is that in view of Theorem 13.6.1 the winding number of the asymptotic eigenvalue satisfies $w(\eta_u) \geq 1$ and as a consequence of the monotonicity of the winding number from Theorem 11.3.2 a fast finite energy plane has a fast asymptotic decay. In view of this notion we can rephrase Corollary 13.6.6 as

Corollary 13.6.8. *A non-degenerate finite energy plane $\widetilde{u} = (u, a)$ is fast if and only if $w(\eta_u) = 1$.*

If $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane, then in view of the definition of α in (11.13) and Theorem 11.3.3 the winding number of $\widetilde{u}$ satisfies the inequality

$$w(\eta_u) \leq \alpha = \left\lfloor \frac{\mu_{CZ}(u)}{2} \right\rfloor$$

where for a real number r we abbreviate by

$$\lfloor r \rfloor := \max\{n \in \mathbb{N} : n \leq r\}$$

the integer part of r. Combining this inequality with Theorem 13.6.5, we obtain further the following corollary.

Corollary 13.6.9. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane. Then the winding number of its asymptotic eigenvalue satisfies*

$$1 \leq w(\eta_u) \leq \left\lfloor \frac{\mu_{CZ}(u)}{2} \right\rfloor.$$

This has the further consequence.

Corollary 13.6.10. *Assume that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane such that $\mu_{CZ}(u) \in \{2, 3\}$. Then $\widetilde{u}$ is fast.*

Chapter 14

Siefring's Intersection Theory for Fast Finite Energy Planes

14.1 Positivity of intersection for closed curves

Suppose that (M, J) is a $2n$-dimensional almost complex manifold, i.e., $J \in \mathrm{End}(TM)$ is an almost complex structure which means that $J^2 = -\mathrm{id}$. We define an orientation on M as follows. If $x \in M$, we declare a basis of T_xM the form $\{v_1, Jv_1, v_2, Jv_2, \dots, v_n, Jv_n\}$ to be positive. Equivalently, that means that the basis $\{v_1, v_2, \dots, v_n, Jv_1, Jv_2, \dots Jv_n\}$ is positive if $\frac{n(n-1)}{2}$ is even and negative otherwise. It remains to explain that this notion is well defined, i.e., independent of the choice of $\{v_1, \dots, v_n\}$. To see that suppose that we have another basis $\{v_1', Jv_1', v_2', Jv_2', \dots, v_n', Jv_n'\}$ of this form. Let $B \in GL(\mathbb{C}^n)$ be the basis change matrix from the complex basis $\{v_1, \dots, v_n\}$ to the complex basis $\{v_1', \dots, v_n'\}$. Write

$$B = A_1 + iA_2$$

where A_1 and A_2 are real $n \times n$-matrices. It follows that the real $2n \times 2n$-matrix

$$A = \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix}$$

is the real basis change matrix from the basis $\{v_1, \dots, v_n, Jv_1, \dots, Jv_n\}$ to the basis $\{v_1', \dots, v_n', Jv_1', \dots, Jv_n'\}$. The determinant of A satisfies

$$\begin{aligned}
\det(A) &= \det \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix} \\
&= \det\left(\begin{pmatrix} \mathrm{id} & \mathrm{id} \\ i\cdot\mathrm{id} & -i\cdot\mathrm{id} \end{pmatrix}^{-1} \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix} \begin{pmatrix} \mathrm{id} & \mathrm{id} \\ i\cdot\mathrm{id} & -i\cdot\mathrm{id} \end{pmatrix} \right) \\
&= \det\left(\frac{1}{2} \begin{pmatrix} \mathrm{id} & -i\cdot\mathrm{id} \\ \mathrm{id} & i\cdot\mathrm{id} \end{pmatrix}^{-1} \begin{pmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{pmatrix} \begin{pmatrix} \mathrm{id} & \mathrm{id} \\ i\cdot\mathrm{id} & -i\cdot\mathrm{id} \end{pmatrix} \right)
\end{aligned}$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_14

$$\begin{aligned}
&= \det\begin{pmatrix} A_1 - iA_2 & 0 \\ 0 & A_1 + iA_2 \end{pmatrix} \\
&= \det(\overline{B}) \cdot \det(B) \\
&= |\det(B)|^2 \\
&> 0.
\end{aligned}$$

This proves that the basis $\{v_1, \ldots, v_n, Jv_1, \ldots, Jv_n\}$ has the same sign as the basis $\{v_1', \ldots, v_n', Jv_1', \ldots, Jv_n'\}$. Consequently, the two bases $\{v_1, Jv_1, \ldots, v_n, Jv_n\}$ and $\{v_1', Jv_1', \ldots, v_n', Jv_n'\}$ have the same sign as well. This shows that the orientation of the almost complex manifold (M, J) is well defined. In the following we always endow an almost complex manifold with this orientation.

Now assume that $M = M^4$ is a four-dimensional almost complex manifold and (Σ_1, i) and (Σ_2, i) are two closed Riemann surfaces. Suppose that

$$u_1 \colon \Sigma_1 \to M^4, \quad u_2 \colon \Sigma_2 \to M^4$$

are two holomorphic maps, i.e.,

$$du_k \circ i = J du_k, \quad k \in \{1, 2\}.$$

We further assume that

$$u_1 \pitchfork u_2,$$

i.e., the two curves intersect transversely, meaning that if $(z_1, z_2) \in \Sigma_1 \times \Sigma_2$ is such that $u_1(z_1) = u_2(z_2)$ it holds that

$$\operatorname{im} du_1(z_1) + \operatorname{im} du_2(z_2) = T_{u_1(z_1)}M = T_{u_2(z_2)}M.$$

Because M is four-dimensional and Σ_1 and Σ_2 are two-dimensional, this is equivalent to requiring

$$\operatorname{im} du_1(z_1) \oplus \operatorname{im} du_2(z_2) = T_{u_1(z_1)}M.$$

We now compute the intersection index at the intersection point $(z_1, z_2) \in \Sigma_1 \times \Sigma_2$. Note that since Σ_1, Σ_2, and M are all almost complex manifolds, they are canonically oriented. Choose a positive basis $\{v_1, iv_1\}$ of $T_{z_1}\Sigma_1$ and a positive basis $\{v_2, iv_2\}$ of $T_{z_2}\Sigma_2$. Using that u_1 and u_2 are holomorphic we get that

$$\begin{aligned}
&\{du_1(z_1)v_1, du_1(z_1)(iv_1), du_2(z_2)v_2, du_2(z_2)(iv_2)\} \\
&\quad = \{du_1(z_1)v_1, Jdu_1(z_1)v_1, du_2(z_2)v_2, Jdu_2(z_2)v_2\}
\end{aligned}$$

is a positive basis of $T_{u_1(z_1)}M$. Therefore the intersection index equals one for every intersection point. In particular, the intersection number of u_1 and u_2 is nonnegative and it vanishes if and only if there are no intersection points. This phenomenon is referred to as *positivity of intersection*. It is a nontrivial fact due to McDuff and Micallef–White that positivity of intersection continues to hold for perturbations of non-transverse intersection points. This is explained in [173, Appendix E]. That means even without the assumption that u_1 and u_2 intersect transversely their algebraic intersection number is still nonnegative and it vanishes if and only if there are no intersection points.

14.2 The algebraic intersection number for finite energy planes

In the following we consider a closed three-dimensional contact manifold (Σ, λ) and fix an SFT-like almost complex structure J on $\Sigma \times \mathbb{R}$. We assume that γ is a non-degenerate Reeb orbit and $\widetilde{u}$ and $\widetilde{v}$ are two finite energy planes whose common asymptotic orbit is γ and $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$. In this section we explain how to associate to $\widetilde{u}$ and $\widetilde{v}$ an algebraic intersection number

$$\mathrm{int}(\widetilde{u}, \widetilde{v}) \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}.$$

The reason that $\mathrm{int}(\widetilde{u}, \widetilde{v})$ is nonnegative is because of positivity of intersection for the two holomorphic curves. Even if $\widetilde{u}$ and $\widetilde{v}$ intersect transversely it is a priori not obvious that the number of intersection points is finite, since the domain $\mathbb{C}$ of $\widetilde{u}$ and $\widetilde{v}$ is not compact. The crucial ingredient which guarantees finiteness of the algebraic intersection number is the following theorem due to Siefring [219].

Theorem 14.2.1. *Suppose that γ is a non-degenerate periodic Reeb orbit which is the common asymptotic limit of two finite energy planes $\widetilde{u}$ and $\widetilde{v}$. Assume that $U, V \colon [R, \infty) \times S^1 \to \gamma^*\xi$ are two asymptotic representatives of $\widetilde{u}$ and $\widetilde{v}$. If $U \neq V$ then there exists $\eta \in \mathfrak{S}(A_\gamma) \cap (-\infty, 0)$ and ζ an eigenvector of A_γ for the eigenvalue η satisfying*

$$V(s,t) - U(s,t) = e^{\eta s}\big(\zeta(t) + \kappa(s,t)\big)$$

and there exist constants $M_{i,j}, d > 0$ for $0 \leq i, j < \infty$ such that

$$|\nabla_s^i \nabla_t^j \kappa(s,t)| \leq M_{i,j} e^{-ds}.$$

Corollary 14.2.2. *Assume that $\widetilde{u}$ and $\widetilde{v}$ are two finite energy planes in $\Sigma \times \mathbb{R}$ asymptotic to the same non-degenerate periodic Reeb orbit γ. Suppose that $\widetilde{u} \pitchfork \widetilde{v}$. Then the number of intersection points between $\widetilde{u}$ and $\widetilde{v}$ is finite, i.e.,*

$$\#\big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} < \infty.$$

Proof. We prove the corollary in 3 steps.

Step 1: *There exists a compact subset $K_0 \subset \mathbb{C}$ with the property that*

$$\big\{(z_1, z_2) \in K_0^c \times K_0^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \emptyset$$

where $K^c = \mathbb{C} \setminus K$ denotes the complement of K in $\mathbb{C}$.

The proof of Step 1 is a bit involved since we have to deal with the possibility that the asymptotic periodic orbit γ is multiply covered. We define the covering number of γ as

$$\mathrm{cov}(\gamma) := \max\Big\{k \in \mathbb{N} : \gamma\big(t + \tfrac{1}{k}\big) = \gamma(t),\ t \in S^1\Big\}.$$

In view of uniqueness of a first-order ODE, we conclude that

$$\gamma(t) \neq \gamma(t'), \quad t - t' \notin \frac{1}{\mathrm{cov}(\gamma)}\mathbb{Z}. \tag{14.1}$$

By definition of an asymptotic representative there exists a compact subset $K_0 \subset \mathbb{C}$ such that there exists a bijection between the following two sets

$$\begin{aligned}&\Big\{(z_1, z_2) \in K_0^c \times K_0^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\Big\}\\ &\quad\cong \Big\{\big((s_1,t_1),(s_2,t_2)\big) \in \big([R,\infty) \times S^1\big) \times \big([R,\infty) \times S^1\big) :\\ &\qquad\qquad \big(\gamma(t_1), U(s_1,t_1), s_1\tau\big) = \big(\gamma(t_2), V(s_2,t_2), s_2\tau\big)\Big\}\end{aligned}$$

where $\tau > 0$ is the period of the periodic orbit γ. Assume that $(s_1, t_1) \in [R,\infty) \times S^1$ and $(s_2, t_2) \in [R,\infty) \times S^1$ such that

$$\big(\gamma(t_1), U(s_1,t_1), s_1\tau\big) = \big(\gamma(t_2), V(s_2,t_2), s_2\tau\big).$$

Because $\tau \neq 0$, we conclude that

$$s_1 = s_2$$

and using (14.1) we infer that

$$t_2 = t_1 + \frac{j}{\mathrm{cov}(\gamma)}, \quad 0 \leq j \leq \mathrm{cov}(\gamma) - 1.$$

Therefore

$$U\Big(s_1, t_1\Big) - V\Big(s_1, t_1 + \tfrac{j}{\mathrm{cov}(\gamma)}\Big) = 0. \tag{14.2}$$

For $j \in \{0, \dots, \mathrm{cov}(\gamma) - 1\}$ define

$$\widetilde{v}_j \colon \mathbb{C} \to \Sigma \times \mathbb{R}, \quad z \mapsto \widetilde{v}(e^{2\pi i \frac{j}{\mathrm{cov}(\gamma)}} z).$$

Note that $\widetilde{v}_j$ is a finite energy plane and in view of the definition of $\mathrm{cov}(\gamma)$ the asymptotic orbit of $\widetilde{v}_j$ is γ as well. An asymptotic representative for $\widetilde{v}_j$ is the map

$$V_j \colon [R,\infty) \times S^1 \to \gamma^*\xi, \quad (s,t) \mapsto V\Big(s, t + \tfrac{j}{\mathrm{cov}(\gamma)}\Big).$$

Equation (14.2) can be reinterpreted as

$$U(s_1, t_1) - V_j(s_1, t_1) = 0.$$

Because $\widetilde{u} \pitchfork \widetilde{v}$ we conclude that

$$U \neq V_j.$$

Hence by Theorem 14.2.1 there exists $\eta \in \mathfrak{S}(A_\gamma) \cap (-\infty, 0)$ and ζ an eigenvector of A_γ to the eigenvalue η such that

$$U(s,t) - V_j(s,t) = e^{\eta s}\big(\zeta(t) + \kappa(s,t)\big)$$

where κ decays exponentially with all its derivatives. Because ζ is an eigenvector of A_γ and therefore a solution of a first-order ODE, it follows that

$$\zeta(t) \neq 0, \quad \forall\, t \in S^1.$$

Because κ decays exponentially, we can assume, perhaps after enlarging K_0 and R, that

$$\sup_{(s,t)\in[R,\infty)\times S^1} |\kappa(s,t)| < \min\big\{|\zeta(t)| : t \in S^1\big\}.$$

Hence we can assume that

$$U(s,t) - V_j(s,t) \neq 0, \quad \forall\, (s,t) \in [R,\infty)\times S^1,\ \forall\, 0 \leq j < \mathrm{cov}(\gamma).$$

Therefore

$$\widetilde{u}(z_1) \neq \widetilde{v}(z_2), \quad \forall\, (z_1,z_2) \in K_0^c \times K_0^c.$$

This finishes the proof of Step 1.

Step 2: *There exists a compact subset $K \subset \mathbb{C}$ with the property that*

$$\big\{(z_1,z_2) \in \mathbb{C}\times\mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \big\{(z_1,z_2) \in K \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}.$$

If $\widetilde{u} = (u,a)$ and $\widetilde{v} = (v,b)$ with $u,v\colon \mathbb{C} \to \Sigma$ and $a,b\colon \mathbb{C} \to \mathbb{R}$ and K_0 is as in Step 1, we abbreviate

$$c := \max\big\{a(z), b(z) : z \in K_0\big\}.$$

In view of the asymptotic behavior there exists $K \supset K_0$ compact such that

$$a(z) > c,\ b(z) > c \quad \forall\, z \in K^c.$$

We decompose

$$\begin{aligned}
\big\{(z_1,z_2) \in \mathbb{C}\times\mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} &= \big\{(z_1,z_2) \in K^c \times K^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&\quad \cup \big\{(z_1,z_2) \in K^c \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&\quad \cup \big\{(z_1,z_2) \in K \times K^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&\quad \cup \big\{(z_1,z_2) \in K \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&=: A_{11} \cup A_{12} \cup A_{21} \cup A_{22}.
\end{aligned}$$

Since $K_0 \subset K$ we have $K^c \subset K_0^c$ and therefore by Step 1

$$A_{11} \subset \big\{(z_1,z_2) \in K_0^c \times K_0^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \emptyset.$$

We decompose the next two sets in the decomposition further, namely

$$\begin{aligned}
A_{12} &= \big\{(z_1,z_2) \in K^c \times K_0 : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&\quad \cup \big\{(z_1,z_2) \in K^c \times K \setminus K_0 : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&\subset \big\{(z_1,z_2) \in K^c \times K_0 : a(z_1) = b(z_2)\big\}\\
&\quad \cup \big\{(z_1,z_2) \in K_0^c \times K_0^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}\\
&= \emptyset
\end{aligned}$$

and similarly we obtain

$$A_{21} = \emptyset.$$

This finishes the proof of Step 2.

Step 3: *We prove the corollary.*

Since $\widetilde{u}$ and $\widetilde{v}$ intersect transversely, the number of intersection points of $\widetilde{u}|_K$ and $\widetilde{v}|_K$ is finite for every compact set K. The corollary now follows immediately from Step 2. □

Remark 14.2.3. In Corollary 14.2.2 we assumed that the two planes approach the same orbit. In fact the assertion of the corollary still holds, when the two planes approach coverings of the same orbit, possibly with different covering numbers. This case can be reduced to the case handled above by precomposing the planes with maps of the form $z \mapsto z^k$ for appropriate choices of k.

In view of Corollary 14.2.2 if γ is a non-degenerate periodic Reeb orbit and $\widetilde{u}$ and $\widetilde{v}$ are two finite energy planes with common asymptotic γ which intersect transversely we define the algebraic intersection number of $\widetilde{u}$ and $\widetilde{v}$ as

$$\mathrm{int}(\widetilde{u}, \widetilde{v}) := \#\big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} \in \mathbb{N}_0. \tag{14.3}$$

If $\widetilde{u}$ and $\widetilde{v}$ do not intersect transversely however satisfy

$$\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$$

we can still define the algebraic intersection number between $\widetilde{u}$ and $\widetilde{v}$ as follows. Because $\widetilde{u}$ and $\widetilde{v}$ have different image the arguments in the proof of Corollary 14.2.2 together with the fact that the intersection points of two holomorphic curves with different image are isolated (see [173, Appendix E]) imply that there exists a compact subset $K \subset \mathbb{C}$ such that if $\widetilde{u}(z_1) = \widetilde{v}(z_2)$ we necessarily have $(z_1, z_2) \in K \times K$. Now perturb $\widetilde{u}$ to $\widetilde{u}'$ and $\widetilde{v}$ to $\widetilde{v}'$ such that there exists a compact subset $K' \subset \mathbb{C}$ such that

$$\widetilde{u}'|_{(K')^c} = \widetilde{u}|_{(K')^c}, \quad \widetilde{v}'|_{(K')^c} = \widetilde{v}|_{(K')^c}, \quad \widetilde{u}' \pitchfork \widetilde{v}'.$$

We do not require that positivity of intersection continues to hold for $\widetilde{u}'$ and $\widetilde{v}'$. We define

$$\nu \colon \big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}'(z_1) = \widetilde{v}'(z_2)\big\} \to \{-1, 1\}$$

by

$$\nu(z_1, z_2) = \begin{cases} 1 & \begin{array}{c}\{\partial_x \widetilde{u}'(z_1), \partial_y \widetilde{u}'(z_1), \partial_x \widetilde{v}'(z_2), \partial_y \widetilde{v}'(z_2)\} \\ \text{positive basis of } T_{\widetilde{u}'(z_1)}(N \times \mathbb{R}) = T_{\widetilde{v}'(z_2)}(N \times \mathbb{R})\end{array} \\ -1 & \text{else.} \end{cases}$$

The algebraic intersection number of $\widetilde{u}$ and $\widetilde{v}$ is defined as

$$\mathrm{int}(\widetilde{u}, \widetilde{v}) = \sum_{\substack{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} \\ \widetilde{u}'(z_1) = \widetilde{v}'(z_2)}} \nu(z_1, z_2) \in \mathbb{Z}. \tag{14.4}$$

By homotopy invariance the algebraic intersection number is independent of the choice of the perturbations $\widetilde{u}'$ and $\widetilde{v}'$, see [177]. Moreover, if $\widetilde{u}$ and $\widetilde{v}$ intersect transversely, then we do not need to perturb the two maps and since the two maps are holomorphic, positivity of intersection implies that (14.4) coincides with (14.3) in this case. In particular, the algebraic intersection number is nonnegative if $\widetilde{u} \pitchfork \widetilde{v}$. It is a nontrivial fact that this continues to hold if $\widetilde{u}$ and $\widetilde{v}$ do not intersect transversely but have just different images. In fact we can choose the perturbations $\widetilde{u}'$ and $\widetilde{v}'$ such that $\nu(z_1, z_2) = 1$ for every intersection point $(z_1, z_2) \in \mathbb{C} \times \mathbb{C}$ of $\widetilde{u}'$ and $\widetilde{v}'$. This is due to results of McDuff and Micallef–White, see [173, Appendix E]. In particular, we have the following theorem.

Theorem 14.2.4. *Assume that $\widetilde{u}$ and $\widetilde{v}$ are two finite energy planes with common non-degenerate asymptotic orbit γ such that* $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$. *Then the algebraic intersection number of $\widetilde{u}$ and $\widetilde{v}$ satisfies*

$$\mathrm{int}(\widetilde{u}, \widetilde{v}) \in \mathbb{N}_0.$$

Moreover, $\mathrm{int}(\widetilde{u}, \widetilde{v}) = 0$ *if and only if $\widetilde{u}$ and $\widetilde{v}$ do not intersect.*

14.3 Siefring's intersection number

Recall that if γ is a periodic Reeb orbit, then the *covering number* of γ is defined as

$$\mathrm{cov}(\gamma) = \max\Big\{k \in \mathbb{N} : \gamma(t + \tfrac{1}{k}) = \gamma(t),\ \forall\, t \in S^1\Big\}.$$

We assume in this section that γ is a non-degenerate periodic orbit which is *simple* in the sense that

$$\mathrm{cov}(\gamma) = 1.$$

We further suppose that $\widetilde{u} = (u, a)$ is a non-degenerate finite energy plane with asymptotic orbit γ.

Choose a trivialization $\mathfrak{T}\colon u^*\xi \to \mathbb{C} \times \mathbb{C}$ which extends to a trivialization $\mathfrak{T}\colon \gamma^*\xi \to S^1 \times \mathbb{C}$. Such a trivialization gives rise to a non-vanishing section $X_{\mathfrak{T}}\colon \mathbb{C} \to u^*\xi$ defined by

$$X_{\mathfrak{T}}(z) = \mathfrak{T}^{-1}_{u(z)} 1 \in \xi_{u(z)}.$$

Since ξ is transverse to the Reeb vector field, there exists $\epsilon_0 = \epsilon_0(\gamma, \mathfrak{T}) > 0$ such that for every $0 < \epsilon \leq \epsilon_0$ it holds that

$$\mathrm{im}(\exp_\gamma \epsilon X_{\mathfrak{T}}) \cap \mathrm{im}\,\gamma = \emptyset$$

where exp is the exponential map with respect to the restriction of the metric $\omega(\cdot, J\cdot)$ to $\Sigma \times \{0\} \subset \Sigma \times \mathbb{R}$. If $\widetilde{u} = (u, a)$ is a finite energy plane with asymptotic orbit γ, we set for $\epsilon \in (0, \epsilon_0]$

$$u_{\mathfrak{T},\epsilon} = \exp_u \epsilon X_{\mathfrak{T}}, \quad \widetilde{u}_{\mathfrak{T},\epsilon} = (u_{\mathfrak{T},\epsilon}, a).$$

Now assume that $\widetilde{u}$ and $\widetilde{v}$ are fast finite energy planes whose asymptotic limit equals the same non-degenerate simple Reeb orbit γ. We define *Siefring's intersection number* for $\widetilde{u}$ and $\widetilde{v}$ as

$$\mathrm{sief}(\widetilde{u}, \widetilde{v}) = \mathrm{int}(\widetilde{u}_{\mathfrak{T},\epsilon}, \widetilde{v}) + 1 \in \mathbb{Z}.$$

This definition is due to Siefring [220] based on previous work by Hutchings [139] and Kriener [154]. Here int denotes the usual algebraic intersection number obtained by the signed count of intersection points of $\widetilde{u}_{\mathfrak{T},\epsilon}$ and $\widetilde{v}$, maybe after a small generic perturbation which makes the two curves intersect transversely. Note that since the curves $\widetilde{u}_{\mathfrak{T},\epsilon}$ and $\widetilde{v}$ have disjoint asymptotics, the number of intersection points is necessarily finite after a small generic perturbation. By homotopy invariance of the algebraic intersection number one observes that $\mathrm{sief}(\widetilde{u}, \widetilde{v})$ is independent of the trivialization $\mathfrak{T}$ and the choice of ϵ. Since the two curves $\widetilde{u}_{\mathfrak{T},\epsilon}$ and $\widetilde{v}$ have different asymptotics, Siefring's intersection number is a homotopy invariant which is not always true for the algebraic intersection number $\mathrm{int}(\widetilde{u}, \widetilde{v})$. Below in Corollary 14.4.6 we will see that, just like the algebraic intersection number, Siefring's intersection number is symmetric, i.e.,

$$\mathrm{sief}(\widetilde{u}, \widetilde{v}) = \mathrm{sief}(\widetilde{v}, \widetilde{u}),$$

although the definition of Siefring's intersection number is not symmetric in the two arguments.

Remark 14.3.1. Siefring's definition is actually more general than the above. He also considers the case when the planes are asymptotic to a multiply covered orbit γ. Siefring's intersection number is in that case defined as

$$\mathrm{sief}(\tilde{u}, \tilde{v}) = \mathrm{int}(\widetilde{u}_{\mathfrak{T},\epsilon}, \widetilde{v}) + \mathrm{cov}(\gamma),$$

where $\mathrm{cov}(\gamma)$ is the covering number of γ. Our applications only need a simply covered orbit, so we will not consider this more general case.

14.4 Siefring's inequality

The following inequality was discovered by Siefring in [220].

Theorem 14.4.1. *Assume that $\widetilde{u}$ and $\widetilde{v}$ are fast finite energy planes asymptotic to the same simple, non-degenerate periodic Reeb orbit γ such that $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$. Then*

$$0 \leq \mathrm{int}(\widetilde{u}, \widetilde{v}) \leq \mathrm{sief}(\widetilde{u}, \widetilde{v}).$$

Before we embark on the proof of Theorem 14.4.1 let us make the following remarks. Since $\widetilde{u}_{\mathfrak{T},\epsilon}$ is not holomorphic anymore, it is far from obvious that Siefring's intersection number turns out to be nonnegative. Later on we are mostly interested in the case where $\mathrm{sief}(\widetilde{u}, \widetilde{v})$ is zero. Then it follows from Siefring's inequality that the algebraic intersection number $\mathrm{int}(\widetilde{u}, \widetilde{v})$ is zero as well, which implies by positivity of intersection that $\widetilde{u}$ and $\widetilde{v}$ do not intersect. Interestingly, Siefring's intersection number is a homotopy invariant. Therefore if it vanishes fast finite energy planes homotopic to $\widetilde{u}$ and $\widetilde{v}$ still do not intersect unless their images coincide.

In [220] Siefring also defined n intersection number for finite energy planes which are not necessarily fast such that the assertion of Theorem 14.4.1 continues to hold. However, in this case one has to add to the algebraic intersection number of $\widetilde{u}_{\mathfrak{T},\epsilon}$ and $\widetilde{v}$ a number bigger than 1. In the proof it will become clear that the reason why one has to add 1 to fast finite energy planes is that 1 coincides with the winding number of the asymptotic eigenvalue of fast finite energy planes.

We first prove a proposition which does not yet require that $\widetilde{u}$ and $\widetilde{v}$ are fast. Recall that if $\widetilde{u}$ and $\widetilde{v}$ have the same non-degenerate asymptotic Reeb orbit γ but different images, then there exists by Theorem 14.2.1 an eigenvalue of the asymptotic operator A_γ such that the difference of asymptotic representatives decays exponentially with weight given by this eigenvalue. Because the asymptotic representatives are unique up to restriction of their domain of definition, this eigenvalue depends only on $\widetilde{u}$ and $\widetilde{v}$ and we abbreviate it by

$$\eta_{\widetilde{u},\widetilde{v}} = \eta_{\widetilde{v},\widetilde{u}} \in \mathfrak{S}(A_\gamma).$$

We can now formulate the proposition as follows.

Proposition 14.4.2. *Assume that γ is a non-degenerate periodic Reeb orbit and $\widetilde{u}$ and $\widetilde{v}$ are two finite energy planes with common simple asymptotic orbit γ such that* $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$. *Then*

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\epsilon}, \widetilde{v}) = \mathrm{int}(\widetilde{u}, \widetilde{v}) - w(\eta_{\widetilde{u},\widetilde{v}}).$$

Proof. In Step 2 of Corollary 14.2.2 we proved that there exists a compact subset $K \subset \mathbb{C}$ with the property that

$$\big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \big\{(z_1, z_2) \in K \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}. \quad (14.5)$$

If $\widetilde{u} = (u, a)$ and $\widetilde{v} = (v, b)$ with $u, v \colon \mathbb{C} \to \Sigma$ and $a, b \colon \mathbb{C} \to \mathbb{R}$ we set

$$c := \max\big\{a(z), b(z) : z \in K\big\}.$$

In view of the asymptotic behavior of non-degenerate finite energy planes there exists a compact set $K_0 \supset K$ such that

$$a(z), b(z) > c, \quad \forall\, z \in K_0^c$$

where K_0^c denotes as usual the complement of K_0 in $\mathbb{C}$. Now choose a third compact subset $K_1 \supset K_0$ with the property that there exists a smooth cutoff function $\beta \in C^\infty(\mathbb{C}, [0,1])$ such that

$$\beta|_{K_0} = 0, \quad \beta|_{K_1^c} = 1.$$

Abbreviate

$$u_{\mathfrak{T},\epsilon,\beta} = \exp_u \beta\epsilon X_{\mathfrak{T}}, \quad \widetilde{u}_{\mathfrak{T},\epsilon,\beta} = (u_{\mathfrak{T},\epsilon,\beta}, a).$$

Note that $\widetilde{u}_{\mathfrak{T},\epsilon,\beta}$ coincides on the complement of the compact subset K_1 with $\widetilde{u}_{\mathfrak{T},\epsilon}$ and therefore

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\epsilon}, \widetilde{v}) = \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}, \widetilde{v}). \tag{14.6}$$

We write the last intersection number as a sum of four intersection numbers

$$\begin{aligned} \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}, \widetilde{v}) \;=\; & \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_{K^c}) + \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_K) \\ & + \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_K, \widetilde{v}|_{K^c}) + \mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_K, \widetilde{v}|_K). \end{aligned} \tag{14.7}$$

In order to compute these four terms we decompose as in the proof of Corollary 14.2.2 the set of intersection points of $\widetilde{u}_{\mathfrak{T},\beta,\epsilon}$ and $\widetilde{v}$ into four disjoint subsets

$$\big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} = A_{11} \cup A_{12} \cup A_{21} \cup A_{22}$$

where

$$\begin{aligned} A_{11} &= \big\{(z_1, z_2) \in K^c \times K^c : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} \\ A_{12} &= \big\{(z_1, z_2) \in K^c \times K : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} \\ A_{21} &= \big\{(z_1, z_2) \in K \times K^c : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} \\ A_{22} &= \big\{(z_1, z_2) \in K \times K : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\}. \end{aligned}$$

We decompose the set A_{12} further as

$$\begin{aligned} A_{12} \;=\; & \big\{(z_1, z_2) \in K_0^c \times K : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} \\ & \cup\big\{(z_1, z_2) \in (K_0 \setminus K) \times K : \widetilde{u}_{\mathfrak{T},\beta,\epsilon}(z_1) = \widetilde{v}(z_2)\big\} \\ =\; & \big\{(z_1, z_2) \in K_0^c \times K : u_{\mathfrak{T},\beta,\epsilon}(z_1) = v(z_2),\ a(z_1) = b(z_2)\big\} \\ & \cup\big\{(z_1, z_2) \in (K_0 \setminus K) \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} \\ \subset\; & \big\{(z_1, z_2) \in K_0^c \times K : a(z_1) = b(z_2)\big\} \\ & \cup\big\{(z_1, z_2) \in K^c \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} \\ =\; & \emptyset. \end{aligned}$$

For the last equality we observed that the second set is empty in view of (14.5) and the first set is empty since on the complement of K_0 the function a takes values bigger than c, whereas on K the function b takes values less than or equal to c by definition of the constant c and the choice of the set K_0. We conclude that

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_K) = 0. \tag{14.8}$$

Since $K \subset K_0$ and β vanishes on K_0 we obtain for the set A_{21}

$$A_{21} = \big\{(z_1, z_2) \in K \times K^c : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \emptyset$$

where we used again (14.5). This implies

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_K, \widetilde{v}|_{K^c}) = 0. \tag{14.9}$$

We next examine the set A_{22}. Again taking advantage that β vanishes on K and using (14.5) we obtain

$$A_{22} = \big\{(z_1, z_2) \in K \times K : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\} = \big\{(z_1, z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z_1) = \widetilde{v}(z_2)\big\}$$

and therefore

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_K, \widetilde{v}|_K) = \mathrm{int}(\widetilde{u}, \widetilde{v}). \tag{14.10}$$

It remains to compute $\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_{K^c})$. Let $U\colon [R,\infty)\times S^1 \to \gamma^*\xi$ be an asymptotic representative for $\widetilde{u}$ and $V\colon [R,\infty)\times S^1 \to \gamma^*\xi$ be an asymptotic representative for $\widetilde{v}$. Because $\widetilde{u}$ and $\widetilde{v}$ have different images, their intersection points are isolated, see [173, Appendix E] and therefore $U \neq V$. It follows from Theorem 14.2.1 that there exists $\eta = \eta_{\widetilde{u},\widetilde{v}} \in \mathfrak{S}(A_\gamma) \cap (-\infty, 0)$ and an eigenvector ζ of A_γ to the eigenvalue η such that

$$U(s,t) - V(s,t) = e^{\eta s}\big(\zeta(t) + \kappa(s,t)\big)$$

where κ decays with all its derivatives exponentially with uniform weight. Therefore maybe after choosing R larger we can assume that for every $s \geq R$ it holds that $U(s,t) - V(s,t) \neq 0$ for every $t \in S^1$ and

$$\deg\left(t \mapsto \frac{\mathfrak{T}\big(U(s,t) - V(s,t)\big)}{\big|\mathfrak{T}\big(U(s,t) - V(s,t)\big)\big|}\right) = w(\mathfrak{T}\zeta) = w(\eta) \tag{14.11}$$

where $w(\eta)$ is the winding of the eigenvalue η as explained in (11.10). Let $\rho \in C^\infty([R,\infty), [0,1])$ be a smooth cutoff function satisfying

$$\rho(s) = \begin{cases} 0 & s \in [R, 2R] \\ 1 & s \in [2R+1, \infty). \end{cases}$$

Define

$$F\colon [R,\infty) \times S^1 \to \mathbb{C}, \quad (s,t) \mapsto \mathfrak{T}\big(U(s,t) - V(s,t)\big) + \rho(s).$$

Abbreviate by

$$Z = [R,\infty) \times S^1$$

the half-infinite cylinder which we identify with the zero section in the bundle $Z \times \mathbb{C}$. Abbreviate further

$$\Gamma_F = \{(z, F(z)) : z \in Z\} \subset Z \times \mathbb{C}$$

the graph of F. Because γ is simple, we get by homotopy invariance of the algebraic intersection number as long as boundaries do not intersect the equality

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_{K^c}) = \mathrm{int}(Z, \Gamma_F). \tag{14.12}$$

If $s \in [R, \infty)$ meets the condition that $F(s,t) \neq 0$ for every $t \in S^1$, we introduce the winding number

$$w_s(F) := \deg\left(t \mapsto \frac{F(s,t)}{|F(s,t)|}\right) \in \mathbb{Z}.$$

Note that, because $U - V$ decays exponentially in the s-variable and ρ is one for s large enough, there exists $S > R$ such that $F(s,t) \neq 0$ for all $s \geq S$ and for all $t \in S^1$. By homotopy invariance we have

$$w_s(F) = w_S(F), \quad s \geq S.$$

By the choice of R and the fact that $\rho(R) = 0$ we further have $F(R,t) \neq 0$ for every $t \in S^1$ and therefore the winding number $w_R(F)$ is well defined. It follows that

$$\mathrm{int}(Z, \Gamma_F) = w_S(F) - w_R(F). \tag{14.13}$$

Using again that $\rho(R) = 0$ we get from (14.11) that

$$w_R(F) = w(\eta). \tag{14.14}$$

Moreover, because $U - V$ decays exponentially, the map $t \mapsto \frac{F(S,t)}{|F(S,t)|}$ is homotopic to the constant map $t \mapsto 1$ and therefore

$$w_S(F) = 0. \tag{14.15}$$

Combining (14.12) and (14.13) with (14.14) and (14.15) we obtain

$$\mathrm{int}(\widetilde{u}_{\mathfrak{T},\beta,\epsilon}|_{K^c}, \widetilde{v}|_{K^c}) = -w(\eta). \tag{14.16}$$

The proposition now follows by combining (14.6) and (14.7) with (14.8), (14.9), (14.11), and (14.16). □

The following lemma enables us to estimate the winding number of $\eta_{\widetilde{u},\widetilde{v}}$ for fast finite energy planes.

Lemma 14.4.3. *Assume that $\widetilde{u}$ and $\widetilde{v}$ are fast finite energy planes with common asymptotic Reeb orbit γ such that $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$.*

(a) *Assume that the asymptotic eigenvectors of $\widetilde{u}$ and $\widetilde{v}$ satisfy $\zeta_{\widetilde{u}} \neq \zeta_{\widetilde{v}}$. Then $w(\eta_{\widetilde{u},\widetilde{v}}) = 1$.*

(b) *In general, $w(\eta_{\widetilde{u},\widetilde{v}}) \leq 1$.*

Proof. We first prove assertion (a). In this case we have

$$\eta_{\widetilde{u},\widetilde{v}} = \max\{\eta_u, \eta_v\}.$$

Since $\widetilde{u}$ and $\widetilde{v}$ are fast, it follows from Corollary 13.6.8 that $w(\eta_{\widetilde{u},\widetilde{v}}) = 1$. This proves assertion (a).

To prove assertion (b) it suffices in view of the already proved assertion (a) to consider the case $\zeta_{\widetilde{u}} = \zeta_{\widetilde{v}}$. In this case we have

$$\eta_{\widetilde{u},\widetilde{v}} < \eta_u = \eta_v.$$

Corollary 11.3.2 tells us that the winding number is monotone and hence using that $\widetilde{u}$ is fast we get

$$w(\eta_{\widetilde{u},\widetilde{v}}) \leq w(\eta_u) = 1.$$

This finishes the proof of the lemma. □

We are now in position to prove Siefring's inequality

Proof of Theorem 14.4.1. The fact that $\mathrm{int}(\widetilde{u},\widetilde{v}) \geq 0$ is a consequence of positivity of intersection for holomorphic curves and was already stated in Theorem 14.2.4. To prove the second inequality we combine the definition of Siefring's intersection number with Proposition 14.4.2 and assertion (b) of Lemma 14.4.3 to get

$$\mathrm{sief}(\widetilde{u},\widetilde{v}) = \mathrm{int}(\widetilde{u}_{\mathfrak{T},\epsilon},\widetilde{v}) + 1 = \mathrm{int}(\widetilde{u},\widetilde{v}) - w(\eta_{\widetilde{u},\widetilde{v}}) + 1 \geq \mathrm{int}(\widetilde{u},\widetilde{v}). \tag{14.17}$$

This finishes the proof of the theorem. □

In the proof of Theorem 14.4.1 we did not use assertion (a) of Lemma 14.4.3. Plugging assertion (a) into (14.17) we see that in "most" cases Siefring's inequality is actually an equality, namely

Theorem 14.4.4. *Assume that $\widetilde{u}$ and $\widetilde{v}$ are fast finite energy planes asymptotic to the same simple, non-degenerate periodic Reeb orbit such that their asymptotic eigenvectors satisfy $\zeta_{\widetilde{u}} \neq \zeta_{\widetilde{v}}$. Then*

$$\mathrm{int}(\widetilde{u},\widetilde{v}) = \mathrm{sief}(\widetilde{u},\widetilde{v}).$$

Recall that $\mathbb{R}$ acts on a fast finite energy plane $\widetilde{u} = (u,a)$ by $r_*(u,a) = (u,a+r)$. Bringing the $\mathbb{R}$-action into play we can give a quantitative statement to what we mean that in "most" cases Siefring's inequality is actually an equality.

Corollary 14.4.5. *Assume that $\widetilde{u} = (u,a)$ and $\widetilde{v} = (v,b)$ are two fast finite energy planes with the same simple asymptotic Reeb orbit γ. Then there exists a subset $C \subset \mathbb{R}$ which is either empty or consists of precisely one point with the property that for every every $r \in \mathbb{R} \setminus C$ the algebraic intersection number $\mathrm{int}(r_*\widetilde{u},\widetilde{v})$ is defined and satisfies*

$$\mathrm{int}(r_*\widetilde{u},\widetilde{v}) = \mathrm{sief}(r_*\widetilde{u},\widetilde{v}) = \mathrm{sief}(\widetilde{u},\widetilde{v}).$$

Proof. That $\mathrm{sief}(r_*\widetilde{u},\widetilde{v}) = \mathrm{sief}(\widetilde{u},\widetilde{v})$ is an immediate consequence of the homotopy invariance of Siefring's intersection number and is true for any $r \in \mathbb{R}$. To prove the first equality we distinguish two cases.

Case 1: *We assume that* $\mathrm{im}(u) \neq \mathrm{im}(v)$.

In this case we have

$$\mathrm{im}(r_*\widetilde{u}) \neq \mathrm{im}(\widetilde{v}), \quad \forall\, r \in \mathbb{R}.$$

Therefore the algebraic intersection number $\mathrm{int}(r_*\widetilde{u}, \widetilde{v})$ is defined for every $r \in \mathbb{R}$. Recall from Lemma 13.5.3 that the asymptotic eigenvector transforms under the $\mathbb{R}$-action as $\zeta_{r_*\widetilde{u}} = e^{-\frac{\eta_u r}{\tau}} \zeta_{\widetilde{u}}$, where $\tau > 0$ is the period of the periodic Reeb orbit γ. Therefore if we define

$$C := \{r \in \mathbb{R} : \zeta_{r_*\widetilde{u}} = \zeta_{\widetilde{v}}\}$$

we conclude with the fact that the asymptotic eigenvector η_u is different from zero that the set C consists of at most one point. If $r \in \mathbb{R} \setminus C$, then Theorem 14.4.4 implies that $\mathrm{int}(r_*\widetilde{u}, \widetilde{v}) = \mathrm{sief}(r_*\widetilde{u}, \widetilde{v})$. This finishes the proof of the corollary for Case 1.

Case 2: *We assume that* $\mathrm{im}(u) = \mathrm{im}(v)$.

We first observe that for a finite energy plane $\widetilde{u} = (u, a) \in \Sigma \times \mathbb{R}$ the map u determines the $\mathbb{R}$-component a up to a constant. Indeed, in view of the holomorphic curve equation the equality

$$u^*\lambda \circ i = da$$

holds true. Therefore because the images of u and v agree, we have

$$\#\{r \in \mathbb{R} : \mathrm{im}(r_*\widetilde{u}) = \mathrm{im}(\widetilde{v})\} = 1, \tag{14.18}$$

i.e., there is a unique $r_0 \in \mathbb{R}$ such that

$$\mathrm{im}((r_0)_*\widetilde{u}) = \mathrm{im}(\widetilde{v}).$$

Because the asymptotic orbit γ is simple, it follows that the asymptotic eigenvectors of $(r_0)_*\widetilde{u}$ and $\widetilde{v}$ agree. Now arguing as in Step 1 we conclude that for every $r \neq r_0$ the asymptotic eigenvectors of $r_*\widetilde{u}$ and $\widetilde{v}$ do not coincide and therefore the algebraic intersection number of $r_*\widetilde{u}$ with $\widetilde{v}$ is well defined and coincides with Siefring's intersection number. This finishes the proof of the corollary. $\square$

Corollary 14.4.6. *Siefring's intersection number is symmetric, i.e.,*

$$\mathrm{sief}(\widetilde{u}, \widetilde{v}) = \mathrm{sief}(\widetilde{v}, \widetilde{u}).$$

Proof. By Corollary 14.4.5 there exists $r \in \mathbb{R}$ such that

$$\mathrm{int}(r_*\widetilde{u}, \widetilde{v}) = \mathrm{sief}(\widetilde{u}, \widetilde{v}), \quad \mathrm{int}((-r)_*\widetilde{v}, \widetilde{u}) = \mathrm{sief}(\widetilde{v}, \widetilde{u}).$$

Note that

$$\mathrm{int}((-r)_*\widetilde{v}, \widetilde{u}) = \mathrm{int}(\widetilde{v}, r_*\widetilde{u}).$$

Therefore in view of the symmetry of the algebraic intersection number we obtain

$$\mathrm{sief}(\widetilde{v}, \widetilde{u}) = \mathrm{int}(\widetilde{v}, r_*\widetilde{u}) = \mathrm{int}(r_*\widetilde{u}, \widetilde{v}) = \mathrm{sief}(\widetilde{u}, \widetilde{v}).$$

This establishes the symmetry of Siefring's intersection number. $\square$

14.5 Computations and applications

Lemma 14.5.1. *Assume that $\widetilde{u} = (u, a)$ is a fast finite energy plane with simple asymptotic Reeb orbit γ with the property that its Siefring self-intersection number does not vanish, i.e.,*

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) \neq 0.$$

Then u intersects the asymptotic orbit γ, i.e.,

$$\mathrm{im}(u) \cap \mathrm{im}(\gamma) \neq \emptyset.$$

Proof. By homotopy invariance of Siefring's intersection number we get

$$\mathrm{sief}(r_* \widetilde{u}, \widetilde{u}) \neq 0, \quad \forall\, r \in \mathbb{R}.$$

It follows from Corollary 14.4.5 that

$$\mathrm{int}(r_* \widetilde{u}, \widetilde{u}) \neq 0, \quad r \in \mathbb{R} \setminus \{0\}.$$

That means for every $r \in \mathbb{R} \setminus \{0\}$ there exist $z_1^r, z_2^r \in \mathbb{C}$ with the property that

$$r_* \widetilde{u}(z_1^r) = \widetilde{u}(z_2^r).$$

Equivalently

$$u(z_1^r) = u(z_2^r),\ a(z_1^r) + r = a(z_2^r), \quad r \in \mathbb{R} \setminus \{0\}. \tag{14.19}$$

In view of the asymptotic behavior the function $a \colon \mathbb{C} \to \mathbb{R}$ is bounded from below. It follows that

$$\lim_{r \to \infty} a(z_2^r) = \infty$$

and therefore

$$\lim_{r \to \infty} |z_2^r| = \infty. \tag{14.20}$$

In view of Step 2 of the proof of Corollary 14.2.2 there exists a compact subset $K \subset \mathbb{C}$ such that

$$z_1^r \in K, \quad r > 0.$$

Hence we can find a subsequence r_ν for $\nu \in \mathbb{N}$ converging to infinity as $\nu \to \infty$ and $z_* \in K$ such that

$$\lim_{\nu \to \infty} z_1^{r_\nu} = z_* \in K.$$

Taking advantage of (14.19) and (14.20) this implies that

$$u(z_*) = \lim_{\nu \to \infty} u(z_1^{r_\nu}) = \lim_{\nu \to \infty} u(z_2^{r_\nu}) \in \mathrm{im}(\gamma).$$

In particular,

$$\mathrm{im}(u) \cap \mathrm{im}(\gamma) \neq \emptyset.$$

The lemma is proved. $\square$

We state the contrapositive of the lemma as an immediate corollary.

Corollary 14.5.2. *Suppose that $\tilde{u} = (u, a)$ is a fast finite energy plane with simple asymptotic Reeb orbit γ. Assume in addition that u does not intersect γ. Then Siefring's self-intersection number of $\widetilde{u}$ vanishes.*

The main idea in the proof of Lemma 14.5.1 was to take advantage of the $\mathbb{R}$-action on finite energy planes. A stronger result can be obtained by letting r go to infinity and interpreting Siefring's self-intersection number of a fast finite energy plane with Siefring's intersection number of the finite energy plane with the orbit cylinder of its asymptotic Reeb orbit. This idea strictly speaking goes beyond the part of Siefring's intersection theory discussed here, since the orbit cylinder is not a finite energy plane anymore but a punctured finite energy plane. The reader is invited to have a look at Siefring's article [220] to see how Siefring's intersection theory works for punctured finite energy planes. Using this technology one obtains the following theorem, which has Lemma 14.5.1 as an immediate corollary.

Theorem 14.5.3. *Assume that $\widetilde{u} = (u, a)$ is a fast finite energy plane with simple asymptotic Reeb orbit γ. Then Siefring's self-intersection number of $\widetilde{u}$ can be computed as*

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) = \#\big\{(z, t) \in \mathbb{C} \times S^1 : u(z) = \gamma(t)\big\}.$$

Proof. Abbreviate by $\tau > 0$ the period of the periodic Reeb orbit γ. For $r \in \mathbb{R}$ define

$$\widetilde{u}_r \colon \mathbb{C} \to \Sigma \times \mathbb{R}$$

by

$$\widetilde{u}_r(e^{2\pi(s+it)}) = \Big(u(e^{2\pi(s+r+it)}), a(e^{2\pi(s+r+it)}) - \tau r\Big), \quad (s, t) \in \mathbb{R} \times S^1$$

and

$$\widetilde{u}_r(0) = \big(u(0), a(0) - \tau r\big) = (-\tau r)_* \widetilde{u}(0).$$

Then in view of the asymptotic behavior of $\widetilde{u}$ the restriction of $\widetilde{u}_r$ to $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ converges in the C^∞_{loc}-topology to the orbit cylinder

$$\widetilde{\gamma} \colon \mathbb{C}^* \to \Sigma \times \mathbb{R}, \quad e^{2\pi(s+it)} \mapsto (\gamma(t), \tau s)$$

as r goes to infinity. In view of the homotopy invariance of Siefring's intersection number we have

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) = \mathrm{sief}(\widetilde{u}_r, \widetilde{u}), \quad \forall\, r \in \mathbb{R}.$$

Hence by letting r go to infinity we obtain

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) = \mathrm{sief}(\widetilde{\gamma}, \widetilde{u}). \tag{14.21}$$

Observe that the map

$$(z, t) \mapsto \Big(z, e^{2\pi\left(\frac{a(z)}{\tau} + it\right)}\Big)$$

gives rise to a bijection between intersection points

$$\big\{(z,t) \in \mathbb{C} \times S^1 : u(z) = \gamma(t)\big\} \cong \big\{(z_1,z_2) \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z_1) = \widetilde{\gamma}(z_2)\big\}.$$

Because $\widetilde{u}$ is fast, u is an immersion transverse to the Reeb vector field R and therefore

$$\widetilde{u} \pitchfork \widetilde{\gamma}.$$

Since both $\widetilde{u}$ and $\widetilde{\gamma}$ are holomorphic, we obtain by positivity of intersection for the algebraic intersection number

$$\mathrm{int}(\widetilde{u},\widetilde{\gamma}) = \#\big\{(z,t) \in \mathbb{C} \times S^1 : u(z) = \gamma(t)\big\}. \tag{14.22}$$

The asymptotic eigenvector $\zeta_{\widetilde{\gamma}}$ of the orbit cylinder vanishes while the asymptotic eigenvector $\zeta_{\widetilde{u}}$ does not vanish. Therefore it follows from Theorem 14.4.4 that

$$\mathrm{int}(\widetilde{u},\widetilde{\gamma}) = \mathrm{sief}(\widetilde{u},\widetilde{\gamma}). \tag{14.23}$$

Combining (14.21), (14.22), and (14.23) the theorem follows. □

One reason why the vanishing of Siefring's self-intersection number of a fast finite energy plane $\widetilde{u} = (u,a)$ is so useful is the fact that it implies that u is an embedding.

Theorem 14.5.4. *Assume that $\widetilde{u} = (u,a)$ is a fast finite energy plane with simple asymptotic orbit γ such that $\mathrm{im}(u) \cap \mathrm{im}(\gamma) = \emptyset$. Then $u\colon \mathbb{C} \to \Sigma$ is an embedding.*

Proof. Since $\widetilde{u}$ is fast, the map u is already an immersion and hence it remains to prove that it is injective. Assume that $z, z' \in \mathbb{C}$ such that

$$u(z) = u(z'). \tag{14.24}$$

We first show that

$$a(z) = a(z'). \tag{14.25}$$

In order to prove (14.25), we argue by contradiction and assume $a(z) \neq a(z')$. Set $c = a(z') - a(z)$. It follows that

$$c_*\widetilde{u}(z) = (u(z), a(z) + c) = (u(z'), a(z')) = \widetilde{u}(z').$$

But since $c \neq 0$, the maps $\widetilde{u}$ and $c_*\widetilde{u}$ have different images. Therefore their algebraic intersection number $\mathrm{int}(\widetilde{u}, c_*\widetilde{u})$ is defined and by positivity of intersection we conclude that

$$\mathrm{int}(\widetilde{u}, c_*\widetilde{u}) \geq 1.$$

Therefore by Theorem 14.4.1 we obtain

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}_c) \geq 1.$$

Since Siefring's intersection number is a homotopy invariant we get

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) \geq 1.$$

However, by assumption of the theorem we have $\mathrm{im}(u) \cap \mathrm{im}(\gamma) = \emptyset$ and therefore by Lemma 14.5.1 it follows that Siefring's self-intersection number of $\widetilde{u}$ vanishes. This contradiction proves (14.25).
It follows from equations (14.24) and (14.25) that

$$\widetilde{u}(z) = \widetilde{u}(z').$$

A point $z \in \mathbb{C}$ is called an *injective point* of $\widetilde{u}$ if and only if $\widetilde{u}^{-1}(\widetilde{u}(z)) = \{z\}$ and $d\widetilde{u}(z) \neq 0$. Abbreviate by

$$I := \{z \in \mathbb{C} : z \text{ injective point of } \widetilde{u}\}$$

the set of injective points and let

$$S := \mathbb{C} \setminus I$$

be its complement, i.e., the set of non-injective points. For a finite energy plane the following alternative holds. Either the set S of non-injective points is discrete, or the finite energy plane is multiply covered in the sense that there exists a finite energy plane $\widetilde{v} \colon \mathbb{C} \to \Sigma \times \mathbb{R}$ and a polynomial p satisfying $\deg(p) \geq 2$ such that

$$\widetilde{u} = \widetilde{v} \circ p.$$

For a proof of this fact we refer to [122, Appendix]. We also recommend looking at the analog of this fact in the closed case, see [173, Proposition 2.5.1,Theorem E.1.2]. We next explain that if a finite energy plane $\widetilde{u} = (u, a)$ is fast, it cannot be multiply covered. To see that suppose by contradiction that there exists a finite energy plane $\widetilde{v} = (v, b)$ such that $\widetilde{u} = \widetilde{v} \circ p$ for a polynomial of degree at least 2. It follows that

$$u = v \circ p.$$

Since the polynomial p has degree at least 2, it must have a critical point, i.e., there exists a point $z \in \mathbb{C}$ such that

$$dp(z) = 0.$$

It follows that

$$du(z) = dv(p(z))dp(z) = 0$$

contradicting the fact that u is an immersion, since it is fast. We have shown that $\widetilde{u}$ cannot be multiply covered and therefore its set S of non-injective points is discrete.

In order to finish the proof of the theorem let us now assume that there exists $z \neq z' \in \mathbb{C}$ such that $\widetilde{u}(z) = \widetilde{u}(z')$. For $\epsilon > 0$ let us abbreviate by $D_\epsilon(z) = \{w \in$

$\mathbb{C} : |w - z| \leq \epsilon\}$ the closed ϵ-ball around z. Because the set S of non-injective points of $\widetilde{u}$ is discrete, we can find $\epsilon > 0$ such that the following condition holds true

$$\mathrm{im}(\widetilde{u}|_{\partial D_\epsilon(z)}) \cap \mathrm{im}(\widetilde{u}|_{\partial D_\epsilon(z')}) = \emptyset.$$

In view of this fact the algebraic intersection number

$$\mathrm{int}(\widetilde{u}|_{D_\epsilon(z)}, \widetilde{u}|_{D_\epsilon(z')}) \in \mathbb{Z}$$

is well defined. Because $\widetilde{u}(z) = \widetilde{u}(z')$ it follows from positivity of intersection that

$$\mathrm{int}(\widetilde{u}|_{D_\epsilon(z)}, \widetilde{u}|_{D_\epsilon(z')}) \geq 1.$$

Choose $c_0 > 0$ such that for every $|c| \leq c_0$ it holds that

$$\mathrm{im}(\widetilde{u}|_{\partial D_\epsilon(z)}) \cap \mathrm{im}(c_*\widetilde{u}|_{\partial D_\epsilon(z')}) = \emptyset.$$

By homotopy invariance of the algebraic intersection number we conclude that

$$\mathrm{int}(\widetilde{u}|_{D_\epsilon(z)}, c_*\widetilde{u}|_{D_\epsilon(z')}) \geq 1.$$

Hence by positivity of intersection if $0 < |c| \leq c_0$ we obtain

$$\mathrm{int}(\widetilde{u}, c_*\widetilde{u}) \geq 1.$$

Again this implies that $\mathrm{sief}(\widetilde{u}, \widetilde{u}) \geq 1$ in contradiction to the assumption of the theorem. This finishes the proof of the theorem. □

We will see later in Lemma 15.6.3 that the converse of Theorem 14.5.4 is true as well. With the help of this lemma we can give now the following characterization of fast finite energy planes with vanishing Siefring self-intersection number.

Theorem 14.5.5. *Assume that $\widetilde{u} = (u, a)$ is a fast finite energy plane with simple asymptotic orbit γ. Then the following assertions are equivalent.*

(i) $\mathrm{sief}(\widetilde{u}, \widetilde{u}) = 0$,

(ii) $\mathrm{im}(u) \cap \mathrm{im}(\gamma) = \emptyset$,

(iii) *u is an embedding,*

(iv) *$\widetilde{u}$ is an embedding.*

Proof. The equivalence of (i) and (ii) follows from Theorem 14.5.3. That (ii) implies (iii) is the content of Theorem 14.5.4. The implication (iii) ⇒ (iv) is obvious. Finally the implication (iv) ⇒ (i) is proved in Lemma 15.6.3 below. □

A further important contribution of Siefring's intersection number is that it can be used to make sure that two fast finite energy planes do not intersect.

Corollary 14.5.6. *Assume that $\widetilde{u} = (u, a)$ is a fast finite energy plane whose asymptotic orbit γ is simple and moreover $\mathrm{im}(u) \cap \mathrm{im}(\gamma) = \emptyset$. Suppose further that $\widetilde{v} = (v, b)$ is a fast finite energy plane which is homotopic to $\widetilde{u}$ and such that $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$. Then*

$$\mathrm{im}(\widetilde{u}) \cap \mathrm{im}(\widetilde{v}) = \emptyset.$$

Proof. By Lemma 14.5.1 we have

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) = 0.$$

Hence by homotopy invariance of Siefring's intersection number we obtain

$$\mathrm{sief}(\widetilde{u}, \widetilde{v}) = 0.$$

Because $\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(\widetilde{v})$ the algebraic intersection number $\mathrm{int}(\widetilde{u}, \widetilde{v})$ is well defined and by Theorem 14.4.1 we have

$$0 \leq \mathrm{int}(\widetilde{u}, \widetilde{v}) \leq \mathrm{sief}(\widetilde{u}, \widetilde{v}) = 0$$

so that we get

$$\mathrm{int}(\widetilde{u}, \widetilde{v}) = 0.$$

By positivity of intersection from Theorem 14.2.4 we conclude that

$$\mathrm{im}(\widetilde{u}) \cap \mathrm{im}(\widetilde{v}) = \emptyset.$$

This finishes the proof of the corollary. □

Corollary 14.5.7. *Under the assumption of Corollary 14.5.6, assume in addition that $\mathrm{im}(u) \neq \mathrm{im}(v)$. Then*

$$\mathrm{im}(u) \cap \mathrm{im}(v) = \emptyset.$$

Proof. Because $\mathrm{im}(u) \neq \mathrm{im}(v)$, it follows that

$$\mathrm{im}(\widetilde{u}) \neq \mathrm{im}(r_* \widetilde{v}), \quad \forall\, r \in \mathbb{R}.$$

Hence it follows from Corollary 14.5.6 that

$$\mathrm{im}(\widetilde{u}) \cap \mathrm{im}(r_* \widetilde{v}) = \emptyset, \quad \forall\, r \in \mathbb{R}.$$

This implies the corollary. □

Chapter 15

The Moduli Space of Fast Finite Energy Planes

15.1 Fredholm operators

Before explaining the class of operators we are interested in, we make some general remarks about Cauchy–Riemann operators. Consider $\mathbb{C}$ with its standard complex structure i but with a measure μ maybe different from the Lebesgue measure, namely

$$\mu = \mu_h := h dx \wedge dy$$

where $h\colon \mathbb{C} \to \mathbb{R}_+$ is a smooth positive function. Abbreviate by

$$\bar{\partial}\colon C^\infty(\mathbb{C}, \mathbb{C}) \to C^\infty(\mathbb{C}, \mathbb{C})$$

the standard Cauchy–Riemann operator on $\mathbb{C}$ which is given for $\zeta \in C^\infty(\mathbb{C}, \mathbb{C})$ by

$$\bar{\partial}\zeta = \partial_x \zeta + i \partial_y \zeta.$$

Then we define the Cauchy–Riemann operator with respect to the measure μ as

$$\bar{\partial}_\mu := \frac{1}{\sqrt{h}} \bar{\partial}\colon C^\infty(\mathbb{C}, \mathbb{C}) \to C^\infty(\mathbb{C}, \mathbb{C}).$$

Of course

$$\ker \bar{\partial}_\mu = \ker \bar{\partial}.$$

The reason why it is natural to consider the operator $\bar{\partial}_\mu$ is that if we endow the target $\mathbb{C}$ with its standard inner product, the norm $|\bar{\partial}_\mu \zeta|$ has an intrinsic description, i.e., a description which only depends on μ, the complex structure i on the domain $\mathbb{C}$, and the inner product on the target $\mathbb{C}$, but not on the coordinates on $\mathbb{C}$ used to define $\bar{\partial}$. Why this is true is explained in the following lemma.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_15

Lemma 15.1.1. *Assume that* $z \in \mathbb{C}$ *and* $\widehat{z} \neq 0 \in T_z\mathbb{C} = \mathbb{C}$ *is an arbitrary non-vanishing tangent vector, then*

$$|\bar{\partial}_\mu \zeta(z)| = \frac{|d\zeta(z)\widehat{z} + id\zeta(z)i\widehat{z}|}{\sqrt{\mu(\widehat{z}, i\widehat{z})}}.$$

Since $\widehat{z}$ *is arbitrary, the right-hand side is intrinsic.*

Proof. For $\widehat{z} \neq 0 \in T_z\mathbb{C}$ we abbreviate

$$f(\widehat{z}) = \frac{|d\zeta(z)\widehat{z} + id\zeta(z)i\widehat{z}|}{\sqrt{\mu(\widehat{z}, i\widehat{z})}}.$$

By definition we have

$$|\bar{\partial}_\mu \zeta(z)| = f(1).$$

Therefore it remains to show that the function f is constant. For $\widehat{z} = \widehat{x} + i\widehat{y} = \widehat{r}\cos\widehat{\theta} + i\widehat{r}\sin\widehat{\theta} = \widehat{r}e^{i\widehat{\theta}} \neq 0 \in T_z\mathbb{C} = \mathbb{C}$ we compute

$$\begin{aligned} d\zeta(z)\widehat{z} + id\zeta(z)i\widehat{z} &= d\zeta(z)\widehat{x} + d\zeta(z)i\widehat{y} + id\zeta(z)i\widehat{x} - id\zeta(z)\widehat{y} \\ &= d\zeta(z)\widehat{x} + id\zeta(x)i\widehat{x} - i\Big(d\zeta(z)\widehat{y} - id\zeta(z)i\widehat{y}\Big) \\ &= (\widehat{x} - i\widehat{y})\Big(d\zeta(z)1 + id\zeta(z)i\Big) \\ &= \widehat{r}e^{-i\widehat{\theta}}\bar{\partial}\zeta(z) \end{aligned}$$

and therefore

$$|d\zeta(z)\widehat{z} + id\zeta(z)i\widehat{z}| = \widehat{r}|\bar{\partial}\zeta(z)|.$$

Moreover,

$$\begin{aligned} \mu(z)(\widehat{z}, i\widehat{z}) &= h(z)(dx \wedge dy)(\widehat{x} + i\widehat{y}, -\widehat{y} + i\widehat{x}) \\ &= h(z)(\widehat{x}^2 + \widehat{y}^2) \\ &= \widehat{r}^2 h(z). \end{aligned}$$

Combining these expressions we conclude that

$$f(\widehat{z}) = \frac{|\bar{\partial}\zeta(z)|}{\sqrt{h(z)}}$$

is independent of $\widehat{z}$. This proves the lemma. $\square$

Choose a function $\gamma \in C^\infty([0,\infty),(0,\infty))$ satisfying

$$\gamma(r) = \begin{cases} 1 & r \leq \frac{1}{2} \\ \frac{1}{4\pi^2 r^2} & r \geq 1. \end{cases}$$

Define the function $h_\gamma \in C^\infty(\mathbb{C}, \mathbb{R}_+)$ by

$$h_\gamma(z) = \gamma(|z|).$$

In the following we endow the complex plane $\mathbb{C}$ with the measure $\mu = \mu_{h_\gamma}$. The measure μ has the following properties. If $D_r = \{z \in \mathbb{C} : |z| \leq r\}$ denotes the ball of radius r centered at the origin, we have

$$\mu|_{D_{\frac{1}{2}}} = dx \wedge dy.$$

Moreover, if $\phi\colon \mathbb{R} \times S^1 \to \mathbb{C} \setminus \{0\}$ is the biholomorphism

$$\phi(s,t) = e^{2\pi(s+it)}, \quad (s,t) \in \mathbb{R} \times S^1,$$

then we obtain

$$\phi^*\mu|_{\mathbb{C}\setminus D_1} = ds \wedge dt.$$

That means on the cylindrical end $\mathbb{C}\setminus D_1$ the measure μ coincides with the standard measure on the cylinder.

Denote by $M_2(\mathbb{R})$ the vector space of real 2×2-matrices. Let

$$S \in C^\infty(\mathbb{C}, M_2(\mathbb{R}))$$

be a smooth family of 2×2-matrices parametrized by $\mathbb{C}$. Suppose that there exists

$$S_\infty \in C^\infty(S^1, \mathrm{Sym}(2))$$

a smooth loop of symmetric 2×2-matrices such that uniformly in the C^∞-topology it holds that

$$\lim_{s \to \infty} 2\pi e^{2\pi s} S(e^{2\pi(s+it)}) = S_\infty(t).$$

We abbreviate by

$$H_1 := W^{1,2}(\mathbb{C}, \mu; \mathbb{C}) \tag{15.1}$$

the Hilbert space of $W^{1,2}$-maps from $\mathbb{C}$ to $\mathbb{C}$ where the domain $\mathbb{C}$ is endowed with the measure μ and the target with the standard inner product and similarly for the L^2-space

$$H_0 := L^2(\mathbb{C}, \mu; \mathbb{C}). \tag{15.2}$$

We consider the bounded linear operator

$$L_S\colon H_1 \to H_0, \quad \zeta \mapsto \frac{1}{\sqrt{h_\gamma}}(\bar{\partial}\zeta + S\zeta) = \bar{\partial}_\mu \zeta + \frac{1}{\sqrt{h_\gamma}} S\zeta. \tag{15.3}$$

From the assumed behavior of S we also get an asymptotic operator

$$A_{S_\infty}: W^{1,2}(S^1, \mathbb{C}) \to L^2(S^1, \mathbb{C}), \quad \zeta \mapsto -J_0 \partial_t \zeta - S_\infty \zeta.$$

Associated to the loop of symmetric matrices S_∞ is the path of symplectic matrices

$$\Psi = \Psi_{S_\infty} \in C^\infty([0,1], \mathrm{Sp}(1))$$

defined by

$$\partial_t \Psi(t) = J_0 S_\infty(t)\Psi(t), \quad \Psi(0) = \mathrm{id}.$$

The following theorem is due to Schwarz [216]

Theorem 15.1.2 (Schwarz). *Assume that* $\ker A_{S_\infty} = \{0\}$, *i.e., the path of symplectic matrices* Ψ_{S_∞} *is non-degenerate. Then* L_S *is a Fredholm operator and its index computes to be*

$$\mathrm{ind} L_S = \mu_{CZ}(\Psi_{S_\infty}) + 1.$$

Remark 15.1.3. The theorem of Schwarz is very reminiscent of the theorem of Riemann–Roch. Indeed, by noting that the Euler characteristic of the complex plane is one, the index formula of Schwarz can be interpreted as

$$\mathrm{ind} L_S = \mu_{CZ}(\Psi_{S_\infty}) + 1 \cdot \chi(\mathbb{C}).$$

We refer to the thesis by Bourgeois [42, Section 5.2] for a derivation of Schwarz' theorem out of the Riemann–Roch formula.

We do not prove Schwarz' theorem, but we illustrate it now for a family of examples.

Illustration of Theorem 15.1.2. Pick a cutoff function $\beta \in C^\infty(\mathbb{R}, [0,1])$ satisfying

$$\beta(s) = \begin{cases} 1 & s \geq 1 \\ 0 & s \leq 0. \end{cases}$$

Choose further

$$\mu \in \mathbb{R} \setminus 2\pi\mathbb{Z}.$$

Define $S = S^\mu \colon \mathbb{C} \to \mathrm{Sym}(2)$ by

$$S(e^{2\pi(s+it)}) = \frac{\beta(s)\mu}{2\pi e^{2\pi s}}\mathrm{id}.$$

We first examine the kernel of the operator L_S. Suppose that $\zeta \in \ker L_S$ and set

$$\eta = \zeta \circ \phi \colon \mathbb{R} \times S^1 \to \mathbb{C}.$$

It follows that η is a solution of the PDE

$$\partial_s \eta + i\partial_t \eta + \beta\mu\eta = 0. \tag{15.4}$$

Write η as a Fourier series

$$\eta(s,t) = \sum_{k=-\infty}^{\infty} \eta_k(s) e^{2\pi i k t}.$$

It follows from (15.4) that each Fourier coefficient is a solution of the ODE

$$\partial_s \eta_k(s) - 2\pi k \eta_k(s) + \beta(s)\mu\eta_k(s) = 0, \quad k \in \mathbb{Z}.$$

Since $\beta(s) = 1$ for $s \geq 1$, we get

$$\eta_k(s) = \eta_k(1)e^{(2\pi k - \mu)(s-1)}, \quad s \geq 1.$$

Because $\eta|_{[0,\infty)\times S^1} \in L^2([0,\infty) \times S^1, \mathbb{C})$, we conclude

$$\eta_k = 0, \quad k > \frac{\mu}{2\pi}\,. \tag{15.5}$$

Since $\beta(s) = 0$ for $s \leq 0$, we conclude that

$$\eta_k(s) = \eta_k(0)e^{2\pi k s}, \quad s \leq 0.$$

Because $\zeta \in W^{1,2}(\mathbb{C}, \mu; \mathbb{C})$, the function is continuous and the limit

$$\lim_{s\to-\infty} \eta(s,t) = \zeta(0)$$

exists. Therefore,

$$\eta_k = 0, \quad k < 0. \tag{15.6}$$

Denoting by $\lfloor \frac{\mu}{2\pi} \rfloor = \max\{n \in \mathbb{Z} : n \leq \frac{\mu}{2\pi}\}$ the integer part of $\frac{\mu}{2\pi}$ we conclude from (15.5) and (15.6) that

$$\dim(\ker L_S) = \left\{ \begin{array}{cc} 2\big(\lfloor \frac{\mu}{2\pi} \rfloor + 1\big) & \lfloor \frac{\mu}{2\pi} \rfloor \geq 0 \\ 0 & \text{else.} \end{array} \right. \tag{15.7}$$

We next compute the dimension of the cokernel of L_S. Suppose that $\eta \in \mathrm{coker} L_S = (\mathrm{im}\, L_S)^{\perp_{H_0}}$. That means that $\eta \in H_0$ satisfies

$$\langle L_S\zeta, \eta\rangle_{H_0} = 0, \quad \forall\, \zeta \in H_1. \tag{15.8}$$

We introduce the function

$$f \in C^\infty(\mathbb{R}, \mathbb{R}_+), \quad s \mapsto 2\pi e^{2\pi s}\sqrt{\gamma(e^{2\pi s})}.$$

Note that f satisfies

$$f(s) = \left\{ \begin{array}{cc} 2\pi e^{2\pi s} & s \leq -\frac{\ln(2)}{2\pi} \\ 1 & s \geq 0. \end{array} \right.$$

We compute using the coordinate change $z = x + iy = e^{2\pi(s+it)}$

$$\begin{aligned} \langle L_S\zeta, \eta\rangle_{H_0} &= \int_{\mathbb{C}} \Big\langle \frac{1}{\sqrt{\gamma(|z|)}}\Big(\partial_x\zeta + i\partial_y\zeta + S\zeta\Big), \eta \Big\rangle \gamma(|z|)dxdy \qquad (15.9) \\ &= \int_{\mathbb{R}}\int_0^1 \Big\langle \frac{1}{2\pi e^{2\pi s}\sqrt{\gamma(e^{2\pi s})}}\Big(\partial_s\zeta + i\partial_t\zeta\Big), \eta \Big\rangle 4\pi^2 e^{4\pi s}\gamma(e^{2\pi s})dsdt \\ &\quad + \int_{\mathbb{R}}\int_0^1 \Big\langle \frac{1}{\sqrt{\gamma(e^{2\pi s})}} S\zeta, \eta \Big\rangle 4\pi^2 e^{4\pi s}\gamma(e^{2\pi s})dsdt \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \int_0^1 \Big\langle \partial_s \zeta + i\partial_t \zeta, \eta \Big\rangle f(s) ds dt + \int_{\mathbb{R}} \int_0^1 \Big\langle \beta\mu\zeta, \eta \Big\rangle f(s)^2 ds dt \\
&= \int_{\mathbb{R}} \int_0^1 \Big\langle \partial_s \zeta + i\partial_t \zeta, \eta \Big\rangle f(s) ds dt + \int_{\mathbb{R}} \int_0^1 \Big\langle \beta\mu\zeta, \eta \Big\rangle f(s) ds dt \\
&= \int_{\mathbb{R}} \int_0^1 \Big\langle \partial_s \zeta + i\partial_t \zeta + \beta\mu\zeta, \eta \Big\rangle f(s) ds dt\,.
\end{aligned}
$$

From (15.8) and (15.9) we obtain

$$
\int_{\mathbb{R}} \int_0^1 \Big\langle \partial_s \zeta + i\partial_t \zeta + \beta\mu\zeta, \eta \Big\rangle f(s) ds dt = 0, \quad \forall\, \zeta \in H_1.
$$

By elliptic regularity for the Cauchy–Riemann equation, this implies that the L^2-function η is smooth and therefore using integration by parts η is a solution of the PDE

$$
-\partial_s \eta + i\partial_t \eta + \left(\mu\beta - \frac{\partial_s f}{f}\right)\eta = 0
$$

or equivalently

$$
-\partial_s \eta + i\partial_t \eta + \big(\mu\beta - \partial_s(\ln f)\big)\eta = 0.
$$

By introducing the function

$$
g \in C^\infty(\mathbb{R}, \mathbb{R}), \quad s \mapsto \mu\beta - \partial_s(\ln f)
$$

this can be written more compactly as

$$
-\partial_s \eta + i\partial_t \eta + g\eta = 0. \tag{15.10}
$$

Note that the function g satisfies

$$
g(s) = \begin{cases} -2\pi & s \leq -1 \\ \mu & s \geq 1. \end{cases}
$$

We again write η as a Fourier series

$$
\eta(s,t) = \sum_{k=-\infty}^{\infty} \eta_k(s) e^{2\pi i k t}.
$$

By (15.10) we obtain for each Fourier coefficient the ODE

$$
-\partial_s \eta_k - 2\pi\eta_k + g\eta_k = 0, \quad k \in \mathbb{Z}.
$$

Since $g(s) = \mu$ for $s \geq 1$, we obtain

$$
\eta_k(s) = \eta_k(1) e^{(\mu - 2\pi k)(s-1)}, \quad s \geq 1.
$$

Because $\eta|_{[0,\infty)\times S^1} \in L^2([0,\infty)\times S^1,\mathbb{C})$, we conclude from this that

$$\eta_k = 0, \quad k \leq \frac{\mu}{2\pi}. \tag{15.11}$$

Using that $g(s) = -2\pi$ for $s \leq -1$ we get

$$\eta_k(s) = \eta_k(-1)e^{-2\pi-2\pi k}s + 1 = \eta_k(-1)e^{-2\pi(k+1)(s+1)}$$

and because η, as a smooth function on $\mathbb{C}$, has to converge when s goes to $-\infty$ we must have

$$\eta_k = 0, \quad k \geq 0. \tag{15.12}$$

Combining (15.11) and (15.12) we get for the dimension of the cokernel of L_S the formula

$$\dim(\mathrm{coker} L_S) = \begin{cases} 2\big(-\lfloor\frac{\mu}{2\pi}\rfloor - 1\big) & \lfloor\frac{\mu}{2\pi}\rfloor < -1 \\ 0 & \text{else.} \end{cases} \tag{15.13}$$

Combining (15.7) and (15.13) we obtain for the index of L_S

$$\mathrm{ind} L_S = \dim(\ker L_S) - \dim(\mathrm{coker} L_S) = 2\bigg(\Big\lfloor\frac{\mu}{2\pi}\Big\rfloor + 1\bigg) = \mu_{CZ}(\Psi_{S_\infty}) + 1.$$

This finishes the illustration of Theorem 15.1.2. $\square$

15.2 The first Chern number

In this section we explain the first Chern number of complex vector bundles over the two-dimensional sphere. Suppose that $E \to S^2$ is a complex vector bundle, satisfying $\mathrm{rk}_{\mathbb{C}} E = n$. We decompose the sphere $S^2 = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$ into the union of the upper and lower hemisphere

$$S^2 = S^2_+ \cup S^2_- \tag{15.14}$$

where

$$S^2_\pm = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : \pm x_3 \geq 0\}.$$

Note that both the upper and the lower hemisphere are diffeomorphic to a closed disk. Since differentiable vector bundles over contractible spaces are trivial, there exist complex trivializations

$$\mathfrak{T}_\pm \colon E|_{S^2_\pm} \to S^2_\pm \times \mathbb{C}^n.$$

By endowing the vector bundle E with a Hermitian metric we can assume that $\mathfrak{T}_\pm$ are actually unitary trivializations. Let

$$S = S^2_+ \cap S^2_- \subset S^2$$

be the equator which we identify with the circle $S^1 = \mathbb{R}/\mathbb{Z}$ via the map $t \mapsto (\cos 2\pi t, \sin 2\pi t, 0)$. If $t \in S$, we obtain a unitary linear map

$$\mathfrak{T}_{+,t}\mathfrak{T}_{-,t}^{-1} \colon \mathbb{C}^n \to \mathbb{C}^n,$$

i.e.,

$$\mathfrak{T}_{+,t}\mathfrak{T}_{-,t}^{-1} \in U(n).$$

We define the first Chern number of E as

$$c_1(E) := \deg\big(t \mapsto \det(\mathfrak{T}_{+,t}\mathfrak{T}_{-,t}^{-1})\big) \in \mathbb{Z}.$$

Note that since $S^2_\pm$ are contractible any two trivializations over the upper or lower hemisphere are homotopic and therefore the first Chern number is independent of the choice of $\mathfrak{T}_\pm$. The first Chern number has the following properties

(i) If $E = S^2 \times \mathbb{C}^n$ is the trivial bundle, it holds that $c_1(E) = 0$.

(ii) If $E_1 \to S^2$ and $E_2 \to S^2$ are two complex vector bundles, the first Chern number of their Whitney sum satisfies

$$c_1(E_1 \oplus E_2) = c_1(E_1) + c_1(E_2) \tag{15.15}$$

and similarly for their tensor product

$$c_1(E_1 \otimes E_2) = c_1(E_1) + c_1(E_2). \tag{15.16}$$

(iii) If $L \to S^2$ is a complex line bundle, then the first Chern class $c_1(L)$ equals the Euler number $e(L)$ of L. In particular, $c_1(L)$ does not depend on the complex structure of L.

Remark 15.2.1. The Euler number of a complex line bundle L over S^2 can be defined as the number of zeros, counted with sign, of a section $\sigma : S^2 \to L$ that intersects the zero-section transversely. This point of view can be found in [41], Theorem 11.16.

To see why (15.15) is true choose unitary trivializations

$$\mathfrak{T}^1_\pm \colon E_1|_{S^2_\pm} \to S^2_\pm \times \mathbb{C}^{n_1}, \quad \mathfrak{T}^2_\pm \colon E_2|_{S^2_\pm} \to S^2_\pm \times \mathbb{C}^{n_2}$$

with $n_1 = \mathrm{rk}_{\mathbb{C}} E_1$ and $n_2 = \mathrm{rk}_{\mathbb{C}} E_2$. We obtain unitary trivializations

$$\mathfrak{T}^1_\pm \oplus \mathfrak{T}^2_\pm \colon (E_1 \oplus E_2)|_{S^2_\pm} \to S^2_\pm \times \mathbb{C}^{n_1+n_2}.$$

For $t \in S^1$ we get

$$\begin{aligned} \det\big((\mathfrak{T}^1_{+,t} \oplus \mathfrak{T}^2_{+,t})(\mathfrak{T}^1_{-,t} \oplus \mathfrak{T}^2_{-,t})^{-1}\big) &= \det\big(\mathfrak{T}^1_{+,t}(\mathfrak{T}^1_{-,t})^{-1} \oplus \mathfrak{T}^2_{+,t}(\mathfrak{T}^2_{-,t})^{-1}\big) \\ &= \det\big(\mathfrak{T}^1_{+,t}(\mathfrak{T}^1_{-,t})^{-1}\big) \cdot \det\big(\mathfrak{T}^2_{+,t}(\mathfrak{T}^2_{-,t})^{-1}\big). \end{aligned}$$

Therefore the first Chern number of the Whitney sum $E_1 \oplus E_2$ satisfies

$$\begin{aligned} c_1(E_1 \oplus E_2) &= \deg\Big(t \mapsto \det\big(\mathfrak{T}^1_{+,t}(\mathfrak{T}^1_{-,t})^{-1}\big) \cdot \det\big(\mathfrak{T}^2_{+,t}(\mathfrak{T}^2_{-,t})^{-1}\big)\Big) \\ &= \deg\Big(t \mapsto \det\big(\mathfrak{T}^1_{+,t}(\mathfrak{T}^1_{-,t})^{-1}\big)\Big) + \deg\Big(t \mapsto \det\big(\mathfrak{T}^2_{+,t}(\mathfrak{T}^2_{-,t})^{-1}\big)\Big) \\ &= c_1(E_1) + c_1(E_2). \end{aligned}$$

This proves (15.15) and (15.16) is proved similarly.

The first Chern number can also be associated to a *symplectic* vector bundle $(E, \omega) \to S^2$, i.e., each fiber (E_z, ω_z) for $z \in S^2$ is a symplectic vector space. In fact the space $\mathcal{J}(E, \omega)$ consisting of all ω-compatible almost complex structures $J \colon E \to E$ is nonempty and contractible. In particular, $\mathcal{J}(E, \omega)$ is connected. Hence we set

$$c_1(E, \omega) := c_1(E, J), \quad J \in \mathcal{J}(E, \omega)$$

and this is well defined by homotopy invariance of the first Chern number, since $\mathcal{J}(E, \omega)$ is connected.

Assume that $E \to S^2$ is a symplectic vector bundle. As usual we identify the equator of S^2 with the circle $S^1 = \mathbb{R}/\mathbb{Z}$. Suppose that for $t \in [0, 1]$ we have a smooth family of symplectic linear maps

$$\Phi_t \colon E_{[0]} \to E_{[t]}$$

such that

$$\Phi_0 = \mathrm{id} \colon E_{[0]} \to E_{[0]}$$

and for $\Phi_1 \colon E_{[0]} \to E_{[1]} = E_{[0]}$ the non-degeneracy condition

$$\ker(\Phi_1 - \mathrm{id}) = \{0\} \tag{15.17}$$

holds true. In this set-up we can associate to the path $t \mapsto \Phi_t$ two Conley–Zehnder indices. Namely, if

$$\mathfrak{T}_+ \colon E|_{S^2_+} \to S^2_+ \times \mathbb{C}^n, \quad 2n = \mathrm{rk}_{\mathbb{R}} E$$

is a symplectic trivialization we define the smooth path

$$\Psi^+ \colon [0, 1] \to \mathrm{Sp}(n), \quad t \mapsto \mathfrak{T}_{+,[t]} \Phi_t \mathfrak{T}^{-1}_{+,[0]}$$

and set

$$\mu^+_{CZ}(\Phi) := \mu_{CZ}(\Psi^+)$$

which is independent of the trivialization $\mathfrak{T}_+$ by homotopy invariance. Similarly we define $\mu^-_{CZ}(\Phi)$ for a symplectic trivialization over S^2_-.

Lemma 15.2.2. *The difference of the two Conley–Zehnder indices satisfies*

$$\mu^+_{CZ}(\Phi) - \mu^-_{CZ}(\Phi) = 2c_1(E).$$

Proof. Choose $J \in \mathcal{J}(E,\omega)$ and choose unitary trivializations $\mathfrak{T}_\pm : E|_{S^2_\pm} \to S^2_\pm \times \mathbb{C}^n$ with respect to J and $\omega(\cdot, J\cdot)$ such that

$$\mathfrak{T}_{+,0} = \mathfrak{T}_{-,0} =: \mathfrak{T}_0 : E_0 \to \mathbb{C}^n.$$

For $t \in [0,1]$ abbreviate

$$U_{[t]} := \mathfrak{T}_{+,[t]}\mathfrak{T}^{-1}_{-,[t]} \in U(n)$$

so that we get a loop

$$U : S^1 \to U(n), \quad [t] \mapsto U_{[t]}.$$

We obtain

$$\Psi^+([t]) = \mathfrak{T}_{+,[t]}\Phi_t\mathfrak{T}^{-1}_{+,[0]} = \mathfrak{T}_{+,[t]}\mathfrak{T}^{-1}_{-,[t]}\mathfrak{T}_{-,[t]}\Phi_t\mathfrak{T}_0^{-1} = U([t])\Psi^-([t]).$$

In particular, the path Ψ^+ is homotopic with fixed endpoints to the concatenation

$$\Psi^+ \cong U\Psi^-(1)\#\Psi^-. \tag{15.18}$$

In view of the non-degeneracy condition (15.17) we have by definition of the Conley–Zehnder index

$$\mu^+_{CZ}(\Phi) = \mu_{CZ}(\Psi^+) = \mu_\Delta(\Gamma_{\Psi^+}) \tag{15.19}$$

where Δ is the diagonal in the symplectic vector space $(\mathbb{C}^n \times \mathbb{C}^n, -\omega \oplus \omega)$ and Γ_{Ψ^+} is the graph of Ψ^+. By homotopy invariance and the concatenation property of the Maslov index we conclude

$$\mu_\Delta(\Gamma_{\Psi_+}) = \mu_\Delta(\Gamma_{\Psi^-}) + \mu_\Delta(\Gamma_{U\Psi^-(1)}).$$

Since U is a loop, we can write this equation as

$$\mu_\Delta(\Gamma_{\Psi_+}) = \mu_{CZ}(\Psi^-) + \mu(\Gamma_U) = \mu^-_{CZ}(\Phi) + \mu(\Gamma_U). \tag{15.20}$$

Since

$$\Gamma_U = \begin{pmatrix} U & 0 \\ 0 & \mathrm{id} \end{pmatrix} \Delta,$$

we compute the Maslov index of the loop Γ_U as

$$\mu(\Gamma_U) = \deg\left(\det\begin{pmatrix} U & 0 \\ 0 & \mathrm{id} \end{pmatrix}^2\right) = \deg\big(\det(U)^2\big) = 2\deg\big(\det(U)\big) = 2c_1(E). \tag{15.21}$$

Combining (15.19), (15.20), and (15.21) the lemma follows. $\square$

15.3 The normal Conley–Zehnder index

Assume that $\widetilde{u}\colon \mathbb{C} \to \Sigma \times \mathbb{R}$ is an *embedded* non-degenerate finite energy plane. Set

$$C := \widetilde{u}(\mathbb{C}) \subset \Sigma \times \mathbb{R}.$$

Since $\widetilde{u}$ is an embedding, the image C is a two-dimensional submanifold of $\Sigma \times \mathbb{R}$. Because $\widetilde{u}$ is holomorphic, the tangent bundle TC is invariant under the SFT-like almost complex structure J. The fiber of the normal bundle

$$NC \to C$$

at a point $c \in C$ is given by $N_c C := (T_c C)^{\perp}$ where the orthogonal complement is taken with respect to the metric $\omega(\cdot, J\cdot)$. Because the almost complex structure J is ω-compatible, the normal bundle NC is invariant under J as well. Because $\mathbb{C}$ is contractible, we can choose a symplectic trivialization

$$\mathfrak{T}_N\colon \widetilde{u}^* NC \to \mathbb{C} \times \mathbb{C}.$$

Moreover, in view of the asymptotic behavior of $\widetilde{u}$ explained in Theorem 13.5.2 we can arrange that $\mathfrak{T}_N$ extends continuously to a smooth symplectic trivialization

$$\mathfrak{T}_N\colon \gamma^*\xi \to S^1 \times \mathbb{C}$$

where γ is the asymptotic Reeb orbit of $\widetilde{u}$. We define the *normal Conley–Zehnder index* of $\widetilde{u}$ as

$$\mu_{CZ}^N(\widetilde{u}) := \mu_{CZ}\Big(t \mapsto \mathfrak{T}_{N,t} d^{\xi}\phi_R^{t\tau}(\gamma(0))\mathfrak{T}_{N,0}^{-1}\Big).$$

Recall from (13.9) that the usual Conley–Zehnder index for $\widetilde{u}$ is defined as follows. Choose a symplectic trivialization

$$\mathfrak{T}_{\xi}\colon u^*\xi \to \mathbb{C} \times \mathbb{C}$$

which extends to a symplectic trivialization

$$\mathfrak{T}_{\xi}\colon \gamma^*\xi \to S^1 \times \mathbb{C}$$

and set

$$\mu_{CZ}(\widetilde{u}) = \mu_{CZ}\Big(t \mapsto \mathfrak{T}_{\xi,t} d^{\xi}\phi_R^{t\tau}(\gamma(0))\mathfrak{T}_{\xi,0}^{-1}\Big).$$

Note that both the usual Conley–Zehnder index and the normal Conley–Zehnder index are independent of the trivializations chosen, because over the contractible space $\mathbb{C}$ all trivializations are homotopic. On the other hand, although $\mathfrak{T}_N$ and $\mathfrak{T}_{\xi}$ trivialize $\gamma^*\xi$, there is no need that the two Conley–Zehnder indices agree, since over the circle two trivializations do not need to be homotopic, in view of the fact that the circle has a nontrivial fundamental group. The following theorem tells us how the two Conley–Zehnder indices are related. It is due to Hofer, Wysocki, and Zehnder, see [128].

Theorem 15.3.1. *Assume $\widetilde{u}$ is an embedded non-degenerate finite energy plane. Then*

$$\mu_{CZ}^N(\widetilde{u}) = \mu_{CZ}(\widetilde{u}) - 2.$$

As preparation for the proof of Theorem 15.3.1 we associate to an embedded non-degenerate finite energy plane $\widetilde{u}$ two complex vector bundles over the two-dimensional sphere S^2

$$\ell_1 \to S^2, \quad \ell_2 \to S^2.$$

Abbreviate by

$$\eta := \langle \partial_r, R \rangle \subset T(\Sigma \times \mathbb{R})$$

the subbundle of $T(\Sigma \times \mathbb{R})$ spanned by the Reeb vector field R and the Liouville vector field ∂_r. Note that this gives rise to a complex splitting of vector bundles

$$T(\mathbb{R} \times \Sigma) = \eta \oplus \xi.$$

We decompose the sphere S^2 into an upper and lower hemisphere, so $S^2 = S^2_+ \cup S^2_-$ as in the discussion about the first Chern number in (15.14). If

$$\mathring{S}^2_\pm = \{(x_1, x_2, x_3) \in S^2 : 0 < \pm x_3 \leq 1\}$$

denotes the interior of $S^2_\pm$, we get diffeomorphisms

$$\psi_\pm \colon \mathring{S}^2_\pm \to \mathbb{C}, \quad (x_1, x_2, x_3) \mapsto e^{2\pi \arctan \frac{\pi\sqrt{x_1^2 + x_2^2}}{2}}(x_1 + ix_2).$$

Note that ψ_+ is orientation preserving, while ψ_- is orientation reversing. In view of the asymptotic behavior of $\widetilde{u}$ explained in Theorem 13.5.2 we get a complex vector bundle ℓ_1 over the sphere S^2 which is characterized by

$$\ell_1|_{\mathring{S}^2_+} = \psi_+^* \widetilde{u}^* TC = \psi_+^* T\mathbb{C}, \quad \ell_1|_{\mathring{S}^2_-} = \psi_-^* \widetilde{u}^* \eta.$$

Note that if we identify the equator $S = S^2_+ \cap S^2_-$ with the circle S^1 by mapping $t \in S^1$ to $(\cos 2\pi t, \sin 2\pi t, 0)$ we obtain

$$\ell_1|_{S^1} = \gamma^* \eta$$

where γ is the asymptotic Reeb orbit of the finite energy plane $\widetilde{u}$. Similarly, we define a complex line bundle ℓ_2 over the sphere S^2 by

$$\ell_2|_{\mathring{S}^2_+} = \psi_+^* \widetilde{u}^* NC, \quad \ell_2|_{\mathring{S}^2_-} = \psi_-^* \widetilde{u}^* \xi.$$

Over the equator this vector bundle satisfies

$$\ell_2|_{S^1} = \gamma^* \xi.$$

Lemma 15.3.2. *The first Chern numbers of the two complex line bundles ℓ_1 and ℓ_2 over S^2 satisfy*

$$1 = c_1(\ell_1) = -c_1(\ell_2).$$

Proof. We define two more vector bundles ℓ_1' and ℓ_2' over the sphere S^2. The vector bundle ℓ_1' is characterized by the conditions

$$\ell_1'|_{\mathring{S}^2_+} = \psi_+^* \widetilde{u}^* \eta, \quad \ell_1'|_{\mathring{S}^2_-} = \psi_-^* \widetilde{u}^* \eta$$

and therefore satisfies over the equator

$$\ell_1'|_{S^1} = \gamma^* \eta.$$

The vector bundle ℓ_2' is determined by

$$\ell_2'|_{\mathring{S}^2_+} = \psi_+^* \widetilde{u}^* \xi, \quad \ell_1'|_{\mathring{S}^2_-} = \psi_-^* \widetilde{u}^* \xi$$

and therefore meets

$$\ell_2'|_{S^1} = \gamma^* \xi.$$

In view of

$$\widetilde{u}^* TC \oplus \widetilde{u}^* NC = \widetilde{u}^* T(\Sigma \times \mathbb{R}) = \widetilde{u}^* \xi \oplus \widetilde{u}^* \eta$$

we obtain

$$\ell_1 \oplus \ell_2 = \ell_1' \oplus \ell_2'.$$

Using the additivity of the first Chern number under Whitney sum (15.15) we get the formula

$$c_1(\ell_1) + c_1(\ell_2) = c_1(\ell_1 \oplus \ell_2) = c_1(\ell_1') + c_1(\ell_2'). \tag{15.22}$$

We next claim that

$$c_1(\ell_1') = c_1(\ell_2') = 0. \tag{15.23}$$

To see that $c_1(\ell_2') = 0$, we first choose a non-vanishing section $\sigma\colon S^1 \to \gamma^* \xi$. We extend σ to a transverse section $\sigma\colon \mathbb{C} \to \widetilde{u}^* \xi$. We define a section $s \in \Gamma(\ell_2')$ by

$$s(x) = \begin{cases} \sigma(\psi_+(x)) & x \in \mathring{S}^2_+ \\ \sigma(\psi_-(x)) & x \in \mathring{S}^2_- \\ \sigma(t) & t \in S^1. \end{cases}$$

By construction all of the zeros of s lie in $\mathring{S}^2_+ \cup \mathring{S}^2_-$ and they appear in pairs. Namely to each zero x of s in $\mathring{S}^2_+$ corresponds the zero $\psi_-^{-1}\psi_+(x) \in \mathring{S}^2_-$ and vice versa. Because ψ_+ is orientation preserving and ψ_- is orientation reversing, the signs of the two zeros cancel and therefore the Euler number of ℓ_2' vanishes. Because the first Chern number of a complex line bundle is just its Euler number we have shown that $c_1(\ell_2') = 0$. The same argument proves that $c_1(\ell_1') = 0$ as well. However, the vanishing of $c_1(\ell_1')$ can be understood even more easily by noting that the Reeb and Liouville vector fields R and ∂_r give rise to a trivialization of ℓ_1'. Formula (15.23) is established.

Combining (15.22) and (15.23) we conclude that

$$c_1(\ell_1) = -c_1(\ell_2) \tag{15.24}$$

and it suffices to show that $c_1(\ell_1) = 1$. To see that, we construct again a section $s \in \Gamma(\ell_1)$ in order to compute the Euler number of ℓ_1. Since $\ell_1|_{\dot{S}^2_-} = \psi_-^* \widetilde{u}^* \eta$, we define $s|_{S^2_-}$ as the pullback of the Liouville vector field ∂_r. Note that therefore $s|_{S^2_-}$ has no zeros. Using that $\widetilde{u}^* TC = T\mathbb{C}$ we have to find for the extension to the upper hemisphere a vector field on $\mathbb{C}$ which points outward asymptotically. For example the vector field $x\partial_x + y\partial_y$ meets this requirement. Note that this vector field has one positive zero at the origin. Since ψ_+ is orientation preserving, we conclude that there exists a section $s \in \Gamma(\ell_1)$ which has precisely one positive zero. That means that the Euler number of ℓ_1 is one and therefore

$$c_1(\ell_1) = e(\ell_1) = 1. \tag{15.25}$$

Equations (15.22) and (15.23) prove the lemma. □

Proof of Theorem 15.3.1. By construction of ℓ_2 we have by Lemma 15.2.2

$$\mu^N_{CZ}(\widetilde{u}) - \mu_{CZ}(\widetilde{u}) = 2c_1(\ell_2).$$

The theorem now follows from Lemma 15.3.2. □

15.4 An implicit function theorem

Assume that γ is a non-degenerate periodic Reeb orbit. Recall that by $\widehat{\mathcal{M}}(\gamma)$ we denote the moduli space of (parametrized) finite energy planes with unparametrized asymptotic orbit $[\gamma]$. The group of direct similitudes $\mathscr{S} = \mathbb{C}^* \ltimes \mathbb{C}$ acts on $\widehat{\mathcal{M}}(\gamma)$ by reparametrization and its quotient $\mathcal{M}(\gamma) = \widehat{\mathcal{M}}(\gamma)/\mathscr{S}$ is the moduli space of unparametrized finite energy planes asymptotic to $[\gamma]$. We denote by

$$\Pi \colon \widehat{\mathcal{M}}(\gamma) \to \mathcal{M}(\gamma), \quad \widetilde{u} \to [\widetilde{u}]$$

the projection. Suppose that $\widetilde{u} \in \widehat{\mathcal{M}}(\gamma)$ is embedded and set $C = \widetilde{u}(\mathbb{C})$. Choose a unitary trivialization $\mathfrak{T} = \mathfrak{T}_N \colon NC \to \mathbb{C} \times \mathbb{C}$, i.e., a trivialization which is complex with respect to the SFT-like almost complex structure J and orthogonal with respect to the inner product $\omega(\cdot, J\cdot)$ and such that it extends to a trivialization $\mathfrak{T} \colon \gamma^*\xi \to S^1 \times \mathbb{C}$. Define a smooth path

$$\Psi = \Psi^{\mathfrak{T}} \colon [0,1] \to \mathrm{Sp}(1)$$

by setting for $t \in [0,1]$

$$\Psi(t) = \mathfrak{T}_t d^\xi \phi_R^{t\tau}(\gamma(0))\mathfrak{T}_0^{-1}.$$

Note that $\Psi(0) = \mathrm{id}\colon \xi_{\gamma(0)} \to \xi_{\gamma(0)}$ and Ψ is non-degenerate in the sense that $\ker(\Psi(1) - \mathrm{id}) = \{0\}$, because γ is non-degenerate. Let

$$S_\infty = S_\infty^{\mathfrak{T}} \in C^\infty([0,1], \mathrm{Sym}(2))$$

generate the path Ψ by

$$\partial_t \Psi = J_0 S_\infty \Psi.$$

In [128] Hofer, Wysocki, and Zehnder construct a smooth map

$$S = S_{\widetilde{u}}^{\mathfrak{T}} \colon \mathbb{C} \to M_2(\mathbb{R}) \tag{15.26}$$

which has the property that $2\pi e^{2\pi s} S(e^{2\pi(s+it)})$ uniformly converges in the C^∞-topology to $S_\infty(t)$, when s goes to infinity. Recall the Hilbert spaces H_1 and H_0 from (15.1) and (15.2), respectively. Then the map S gives rise to a bounded linear operator

$$L_S \colon H_1 \to H_0$$

as defined in (15.3). One should think of the operator L_S as the linearization of the holomorphic curve equation modulo the action of the group of direct similitudes $\mathscr{S}$ by reparametrization. That we mod out the reparametrization action is due to the fact that we only consider variations in the normal direction. The importance of the operator L_S lies in the following implicit function theorem proved by Hofer, Wysocki, and Zehnder in [128].

Theorem 15.4.1. *Assume that $\widetilde{u} \in \widehat{\mathcal{M}}(\gamma)$ is embedded and the operator $L_S \colon H_1 \to H_0$ is surjective. Then locally around $[\widetilde{u}]$ the space $\mathcal{M}(\gamma)$ is a manifold and there exists $\mathcal{U} \subset \ker(L_S)$, an open neighborhood of 0, and a map*

$$\widehat{\mathcal{F}} \colon \mathcal{U} \to \widehat{\mathcal{M}}(\gamma)$$

satisfying $\widehat{\mathcal{F}}(0) = \widetilde{u}$, such that the map

$$\mathcal{F} = \Pi\widehat{\mathcal{F}} \colon \mathcal{U} \to \mathcal{M}(\gamma)$$

defines a local chart around $[\widehat{u}]$. Moreover, if $\zeta \in \mathcal{U}$, the intersection points of $\widetilde{u}$ and $\widehat{\mathcal{F}}(\zeta)$ are related by

$$\big\{(z,z') \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z) = \widehat{\mathcal{F}}(\zeta)(z')\big\} = \big\{(z,z) : z \in \mathbb{C},\ \zeta(z) = 0\big\}.$$

Using Theorem 15.1.2 and Theorem 15.3.1 we can under the assumptions of Theorem 15.4.1 compute the local dimension of the moduli space $\mathcal{M}(\gamma)$ at $[\widetilde{u}]$ by

$$\begin{aligned} \dim_{[\widetilde{u}]} \mathcal{M}(\gamma) &= \dim \ker L_S \\ &= \mathrm{ind} L_S = \mu_{CZ}(\Psi_{S_\infty}) + 1 \\ &= \mu_{CZ}^N(\widetilde{u}) + 1 \\ &= \mu_{CZ}(\widetilde{u}) - 1. \end{aligned} \tag{15.27}$$

15.5 Exponential weights

Suppose that γ is a non-degenerate Reeb orbit. Abbreviate by

$$\widehat{\mathcal{M}}_{\mathrm{fast}}(\gamma) \subset \widehat{\mathcal{M}}(\gamma)$$

the subspace of fast finite energy planes asymptotic to γ. Assume that $\widetilde{u} = (u, a) \in \widehat{\mathcal{M}}_{\mathrm{fast}}(\gamma)$ is embedded. Recall that if $U(s,t) \in \xi_{\gamma(t)}$ is an asymptotic representative of $\widetilde{u}$ we can write

$$U(s,t) = e^{\eta s}\big(\zeta(t) + \kappa(s,t)\big)$$

where κ decays with all its derivatives exponentially, $\eta = \eta_u \in \mathfrak{S}(A_\gamma) \cap (-\infty, 0)$ is a negative eigenvalue of the asymptotic operator and $\zeta = \zeta_{\widetilde{u}} \in \Gamma^{1,2}(\gamma^*\xi)$ is an eigenvector of the asymptotic operator A_γ to the eigenvalue η. Pick further a unitary trivialization

$$\mathfrak{T}_\xi \colon u^*\xi \to \mathbb{C} \times \mathbb{C}$$

which extends to a trivialization

$$\mathfrak{T}_\xi \colon \gamma^*\xi \to S^1 \times \mathbb{C}.$$

In particular, $\mathfrak{T}_\xi\zeta \in C^\infty(S^1, \mathbb{C})$ and the winding of the eigenvalue η is defined as

$$w(\eta) = w(\mathfrak{T}_\xi\zeta) = \deg\left(t \mapsto \frac{\mathfrak{T}_\xi\zeta(t)}{|\mathfrak{T}_\xi\zeta(t)|}\right).$$

Because $\widetilde{u}$ is fast, we have by Corollary 13.6.8

$$w(\eta) = 1.$$

Recall further from Corollary 11.3.2 that the winding number is monotone, i.e., if $\eta \le \eta'$, then $w(\eta) \le w(\eta')$. If $\mu_{CZ}(u) \ge 3$, choose $\delta \le 0$ such that

$$\max\big\{\eta \in \mathfrak{S}(A_\gamma) : w(\eta) = 1\big\} < \delta < \min\big\{\eta \in \mathfrak{S}(A_\gamma) : w(\eta) = 2\big\}.$$

That a non-positive δ with this property exists is guaranteed by Theorem 11.3.3 in view of the assumption that $\mu_{CZ}(u) \ge 3$. If $\mu_{CZ}(u) \le 2$ it follows from Theorem 13.6.1 that $\mu_{CZ}(u) = 2$ and we set in this case $\delta = 0$. Because $\widetilde{u}$ is embedded, we set $C = \widetilde{u}(\mathbb{C}) \in \Sigma \times \mathbb{R}$ and choose in addition a unitary trivialization

$$\mathfrak{T}_N \colon u^*NC \to \mathbb{C} \times \mathbb{C}$$

which extends to a trivialization

$$\mathfrak{T}_N \colon \gamma^*\xi \to S^1 \times \mathbb{C}.$$

This trivialization gives rise to a smooth map

$$S = S_{\widetilde{u}}^{\mathfrak{T}_N} \colon \mathbb{C} \to M_2(\mathbb{R})$$

as mentioned in (15.26). Recall that $\phi\colon \mathbb{R} \times S^1 \to \mathbb{C} \setminus \{0\}$ denotes the biholomorphism $(s,t) \mapsto e^{2\pi(s+it)}$. For $\delta \leq 0$ we introduce the Hilbert space

$$H_1^\delta = \Big\{\zeta \in H_1 : e^{-\delta s}\zeta \circ \phi|_{[0,\infty)\times S^1} \in W^{1,2}\big([0,\infty) \times S^1, \mathbb{C}\big)\Big\},$$

i.e., the subvector space of functions in H_1 introduced in (15.1) which decay on the cylindrical end exponentially with weight $|\delta|$. Similarly, we introduce

$$H_0^\delta = \Big\{\zeta \in H_0 : e^{-\delta s}\zeta \circ \phi|_{[0,\infty)\times S^1} \in L^2\big([0,\infty) \times S^1, \mathbb{C}\big)\Big\}.$$

We define the bounded linear operator

$$L_S^\delta\colon H_1^\delta \to H_0^\delta, \quad \zeta \mapsto \bar{\partial}_\mu \zeta + \frac{1}{\sqrt{h_\gamma}} S\zeta.$$

Note that L_S^δ is given by the same formula as the operator L_S introduced in (15.3), however, its domain and target differ from the ones of L_S. This fact can be expressed with the following commutative diagram

$$\begin{tikzcd} H_1^\delta \arrow[r, "L_S^\delta"] \arrow[d] & H_0^\delta \arrow[d] \\ H_1 \arrow[r, "L_S"] & H_0 \end{tikzcd}$$

where the vertical arrows stand for the inclusion maps.

The analog of the implicit function theorem stated in Theorem 15.4.1 for the fast case can now be formulated as follows.

Theorem 15.5.1. *Assume that $\widetilde{u} \in \widehat{\mathcal{M}}_{\text{fast}}(\gamma)$ is embedded and that the operator $L_S^\delta\colon H_1^\delta \to H_0^\delta$ is surjective. Then locally around $[\widetilde{u}]$ the space $\mathcal{M}_{\text{fast}}(\gamma)$ is a manifold and there exists $\mathcal{U} \subset \ker(L_S^\delta)$, an open neighborhood of 0, and a map*

$$\widehat{\mathcal{F}}\colon \mathcal{U} \to \widehat{\mathcal{M}}_{\text{fast}}(\gamma)$$

satisfying $\widehat{\mathcal{F}}(0) = \widetilde{u}$, such that the map

$$\mathcal{F} = \Pi\widehat{\mathcal{F}}\colon \mathcal{U} \to \mathcal{M}_{\text{fast}}(\gamma)$$

defines a local chart around $[\widehat{u}]$. Moreover, if $\zeta \in \mathcal{U}$, then the intersection points of $\widetilde{u}$ and $\widehat{\mathcal{F}}(\zeta)$ are related by

$$\big\{(z,z') \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z) = \widehat{\mathcal{F}}(\zeta)(z')\big\} = \big\{(z,z) : z \in \mathbb{C},\ \zeta(z) = 0\big\}.$$

Note that if $\widetilde{u}$ is an embedded fast finite energy plane and both $L_S\colon H_1 \to H_0$ and $L_S^\delta\colon H_1^\delta \to H_0^\delta$ are surjective, then locally around $[\widetilde{u}]$ both $\mathcal{M}(\gamma)$ and $\mathcal{M}_{\text{fast}}(\gamma)$ are smooth manifolds and we have

$$\ker(L_S^\delta) = T_{[\widetilde{u}]}\mathcal{M}_{\text{fast}}(\gamma) \subset T_{[\widetilde{u}]}\mathcal{M}(\gamma) = \ker(L_S).$$

Our next goal is to compute the Fredholm index of the operator L_S^δ. The following result is due to Hryniewicz [133].

Theorem 15.5.2. *Assume that $\widetilde{u} = (u, a)$ is an embedded fast finite energy plane. Then the Fredholm index of the Fredholm operator L_S^δ satisfies*

$$\text{ind} L_S^\delta = \begin{cases} 2 & \mu_{CZ}(u) \geq 3 \\ 1 & \mu_{CZ}(u) = 2. \end{cases}$$

Proof. If $\mu_{CZ}(u) = 2$, we have chosen $\delta = 0$ and hence in this case we have $L_S^\delta = L_S$. Hence by Theorem 15.1.2 and Theorem 15.3.1 we obtain

$$\text{ind} L_S^\delta = \mu_{CZ}(\Psi_{S_\infty}) + 1 = \mu_{CZ}^N(u) + 1 = \mu_{CZ}(u) - 1 = 1.$$

Therefore we can assume in the following that

$$\mu_{CZ}(u) \geq 3.$$

In order to compute the Fredholm index of the operator L_S^δ in this case, we first choose a smooth function $\rho\colon \mathbb{R} \to \mathbb{R}$ satisfying

$$\rho(s) = \begin{cases} s & s \geq 1 \\ 0 & s \leq 0. \end{cases}$$

We define a Hilbert space isomorphism

$$T_\delta\colon H_1^\delta \to H_1$$

which is given for $\zeta \in H_1^\delta$ by

$$T_\delta \zeta(e^{2\pi(s+it)}) = e^{-\delta\rho(s)}\zeta(e^{2\pi(s+it)}),\ (s,t) \in \mathbb{R} \times S^1, \quad T_\delta\zeta(0) = \zeta(0).$$

The isomorphism T_δ extends by the same formula to an isomorphism

$$T_\delta\colon H_0^\delta \to H_0.$$

Note that its inverse is given by

$$T_\delta^{-1} = T_{-\delta}\colon H_i \to H_i^\delta, \quad i \in \{0, 1\}.$$

We consider

$$T_\delta L_S^\delta T_{-\delta}\colon H_1 \to H_0.$$

To describe this map we introduce $S_\delta \in C^\infty(\mathbb{C}, M_2(\mathbb{R}))$ by

$$S_\delta(e^{2\pi(s+it)}) := S(e^{2\pi(s+it)}) + \frac{\delta\rho'(s)}{2\pi e^{2\pi s}} \cdot \mathrm{id}, \quad (s,t) \in \mathbb{R} \times S^1.$$

If $\zeta \in H_1$, we compute

$$\begin{aligned}
T_\delta L_S^\delta T_{-\delta}\zeta(e^{2\pi(s+it)}) &= T_\delta L_S^\delta\big(e^{\delta\rho(s)}\zeta(e^{2\pi(s+it)})\big)\\
&= T_\delta \frac{1}{\sqrt{\gamma(e^{2\pi s})}}\big(\partial_x + i\partial_y + S\big)\big(e^{\delta\rho(s)}\zeta(e^{2\pi(s+it)})\big)\\
&= T_\delta \frac{1}{\sqrt{\gamma(e^{2\pi s})}}\left(\frac{1}{2\pi e^{2\pi s}}(\partial_s + i\partial_t) + S\right)\big(e^{\delta\rho(s)}\zeta(e^{2\pi(s+it)})\big)\\
&= T_\delta \frac{e^{\delta\rho(s)}}{\sqrt{\gamma(e^{2\pi s})}}\left(\frac{1}{2\pi e^{2\pi s}}(\partial_s + i\partial_t) + S + \frac{\delta\rho'(s)}{2\pi e^{2\pi s}}\right)\zeta(e^{2\pi(s+it)})\\
&= \frac{1}{\sqrt{\gamma(e^{2\pi s})}}\left(\frac{1}{2\pi e^{2\pi s}}(\partial_s + i\partial_t) + S + \frac{\delta\rho'(s)}{2\pi e^{2\pi s}}\right)\zeta(e^{2\pi(s+it)})\\
&= \frac{1}{\sqrt{\gamma(e^{2\pi s})}}\left(\partial_x + i\partial_y + S + \frac{\delta\rho'(s)}{2\pi e^{2\pi s}}\right)\zeta(e^{2\pi(s+it)})\\
&= \frac{1}{\sqrt{\gamma(e^{2\pi s})}}\big(\partial_x + i\partial_y + S_\delta\big)\zeta(e^{2\pi(s+it)})\\
&= L_{S_\delta}\zeta(e^{2\pi(s+it)}).
\end{aligned}$$

We showed

$$T_\delta L_S^\delta T_{-\delta} = L_{S_\delta}.$$

Note that

$$\begin{aligned}
\lim_{s\to\infty} 2\pi e^{2\pi s} S_\delta(e^{2\pi(s+it)}) &= \lim_{s\to\infty} 2\pi e^{2\pi s} S(e^{2\pi(s+it)}) + \delta\cdot\mathrm{id}\\
&= S_\infty(t) + \delta\cdot\mathrm{id}\\
&=: S_\infty^\delta(t).
\end{aligned}$$

From Theorem 15.1.2 we obtain

$$\mathrm{ind}(L_S^\delta) = \mathrm{ind}(L_{S_\delta}) = \mu_{CZ}(\Psi_{S_\infty^\delta}) + 1. \tag{15.28}$$

It remains to compute the Conley–Zehnder index of the symplectic path $\Psi_{S_\infty^\delta}$. Recall that the linear operator $A_{S_\infty} = -J\partial_t - S_\infty \colon W^{1,2}(S^1,\mathbb{C}) \to L^2(S^1,\mathbb{C})$ equals

$$A_{S_\infty} = A_\gamma^{\mathfrak{T}_N} =: A^N$$

where the operator $A_\gamma^{\mathfrak{T}_N}$ as explained in (13.7) is conjugated to the asymptotic operator $A_\gamma \colon \Gamma^{1,2}(\gamma^*\xi) \to \Gamma^{0,2}(\gamma^*\xi)$ via the trivialization $\mathfrak{T}_N$. Because the winding numbers of the eigenvalues of A_γ are computed with respect to the trivialization $\mathfrak{T}_\xi$, we also need to consider the operator

$$A^\xi := A_\gamma^{\mathfrak{T}_\xi}.$$

Since both operators A^N and A^ξ are conjugated to A_γ they are conjugated to each other as well. To see how A^N and A^ξ are conjugated we consider the loop of unitary transformations $t \mapsto \mathfrak{T}_{N,t}\mathfrak{T}_{\xi,t}^{-1}$ from $\mathbb{C}$ to itself. Because a unitary transformation of $\mathbb{C}$ is just multiplication by a complex number of norm one, we can think of $\mathfrak{T}_{N,t}\mathfrak{T}_{\xi,t}^{-1}$ for each $t \in S^1$ as a unit complex number. Recall from Lemma 15.3.2 that $c_1(\ell_2) = -1$, which implies that

$$\deg\Big(t \mapsto \mathfrak{T}_{N,t}\mathfrak{T}_{\xi,t}^{-1}\Big) = -1.$$

Therefore by replacing the trivializations $\mathfrak{T}_N$ and $\mathfrak{T}_\xi$ by homotopic ones we can assume without loss of generality that

$$\mathfrak{T}_{N,t}\mathfrak{T}_{\xi,t}^{-1} = e^{-2\pi i t}.$$

Consider the Hilbert space isomorphism

$$\Phi\colon W^{1,2}(S^1,\mathbb{C}) \to W^{1,2}(S^1,\mathbb{C})$$

which is given for $v \in W^{1,2}(S^1,\mathbb{C})$ by

$$\Phi(v)(t) = e^{2\pi i t}v(t), \quad t \in S^1$$

with inverse

$$\Phi^{-1}\colon W^{1,2}(S^1,\mathbb{C}) \to W^{1,2}(S^1,\mathbb{C}), \quad \Phi^{-1}v(t) = e^{-2\pi i t}v(t),\ t \in S^1.$$

Note that both Φ and Φ^{-1} extend to Hilbert space isomorphisms

$$\Phi, \Phi^{-1}\colon L^2(S^1,\mathbb{C}) \to L^2(S^1,\mathbb{C}).$$

Note that if $v \in W^{1,2}(S^1,\mathbb{C})$ we have

$$A^N v = \mathfrak{T}_N A_\gamma \mathfrak{T}_N^{-1} v = \mathfrak{T}_N \mathfrak{T}_\xi^{-1} A^\xi \mathfrak{T}_\xi \mathfrak{T}_N^{-1} v = \mathfrak{T}_N \mathfrak{T}_\xi^{-1} A^\xi (\mathfrak{T}_N \mathfrak{T}_\xi)^{-1} v$$

so that we obtain

$$A^N = \Phi^{-1} A^\xi \Phi.$$

Now if $\eta \in \mathfrak{S}(A^\xi)$ is an eigenvalue of A^ξ and v is an eigenvector of A^ξ to the eigenvalue η, it follows that $\Phi^{-1}v$ is an eigenvector of A^N to the eigenvalue η. However, note that the winding number changes under this transformation, namely

$$w(\Phi^{-1}v) = w(v) - 1.$$

Therefore, even though the spectra of the conjugated operators A^ξ and A^N agree, if $\eta \in \mathfrak{S}(A^\xi) = \mathfrak{S}(A^N)$ the corresponding winding numbers differ by

$$w(\eta, A^N) = w(\eta, A^\xi) - 1.$$

Because $S_\infty^\delta = S_\infty + \delta$ we have

$$A_{S_\infty^\delta} = A_{S_\infty} - \delta I = A^N - \delta I =: A_\delta^N$$

where $I \colon W^{1,2}(S^1, \mathbb{C}) \to L^2(S^1, \mathbb{C})$ is the inclusion. We further abbreviate

$$A_\delta^\xi := A^\xi - \delta I.$$

Note that

$$A_\delta^N = \Phi^{-1} A_\delta^\xi \Phi.$$

Therefore

$$\begin{aligned} \alpha(S_\infty^\delta) &= \max\Big\{ w(\eta, A_\delta^N) : \eta \in \mathfrak{S}(A_\delta^N) \cap (-\infty, 0) \Big\} \\ &= \max\Big\{ w(\eta, A_\delta^\xi) : \eta \in \mathfrak{S}(A_\delta^\xi) \cap (-\infty, 0) \Big\} - 1. \end{aligned} \tag{15.29}$$

Note that we have a bijection

$$\mathfrak{S}(A^\xi) \cong \mathfrak{S}(A_\delta^\xi), \quad \eta \mapsto \eta - \delta.$$

Indeed, v is an eigenvector of the operator A^ξ to the eigenvalue η if and only if v is an eigenvector of the operator A_δ^ξ to the eigenvalue $\eta - \delta$. In particular,

$$w(\eta, A^\xi) = w(\eta - \delta, A_\delta^\xi).$$

Hence

$$\begin{aligned} &\max\Big\{ w(\eta, A_\delta^\xi) : \eta \in \mathfrak{S}(A_\delta^\xi) \cap (-\infty, 0) \Big\} \\ &= \max\Big\{ w(\eta + \delta, A^\xi) : \eta \in \mathfrak{S}(A_\delta^\xi), \eta < 0 \Big\} \\ &= \max\Big\{ w(\eta, A^\xi) : \eta \in \mathfrak{S}(A_\delta^\xi), \eta < \delta \Big\} \\ &= 1 \end{aligned} \tag{15.30}$$

by the choice of δ. Combining (15.29) with (15.30) we obtain

$$\alpha(S_\infty^\delta) = 1 - 1 = 0. \tag{15.31}$$

Recall that the parity is defined as

$$p(S_\infty^\delta) = \begin{cases} 0 & \exists\, \eta \in \mathfrak{S}(A_\delta^N) \cap [0, \infty) \text{ such that } \alpha(S_\infty^\delta) = w(\eta, S_\infty^\delta) \\ 1 & \text{else.} \end{cases}$$

By the choice of δ we have

$$p(S^\delta_\infty) = 1. \tag{15.32}$$

Combining (15.31) and (15.32) with Theorem 11.3.3 we obtain

$$\mu_{CZ}(\Psi_{S^\delta_\infty}) = 2\alpha(S^\delta_\infty) + p(S^\delta_\infty) = 1. \tag{15.33}$$

In combination with (15.28), Equation (15.33) implies

$$\mathrm{ind}(L^\delta_S) = 2.$$

This finishes the proof of the theorem. $\square$

15.6 Automatic transversality

We first explain the following local version of automatic transversality for fast finite energy planes.

Lemma 15.6.1. *Assume that $\widetilde{u} = (u, a)$ is an embedded fast finite energy plane with asymptotic orbit γ satisfying $\mu_{CZ}(u) \geq 3$. Then locally around $[\widetilde{u}]$ the moduli space $\mathcal{M}_{\mathrm{fast}}(\gamma)$ is a two-dimensional manifold.*

To prove Lemma 15.6.1 we need the following result, see also [133].

Lemma 15.6.2. *Assume that $\mu_{CZ}(\Psi_{S_\infty}) = 1$, then L_S is surjective.*

Proof. Because $\mu_{CZ}(\Psi_{S_\infty}) = 1$ we obtain from Theorem 15.1.2 that

$$\mathrm{ind}(L_S) = 2.$$

Hence it suffices to show that

$$\dim \ker L_S \leq 2.$$

Suppose that $v \neq 0 \in \ker L_S$. In view of the asymptotic behavior there exists $R_0 > 0$ such that for every $R \geq R_0$

$$\deg\left(t \mapsto \frac{v(Re^{2\pi i t})}{|v(Re^{2\pi i t})|}\right) = w(\eta)$$

where $w(\eta)$ is the winding number of a negative eigenvalue $\eta \in \mathfrak{S}(A_{S_\infty}) \cap (-\infty, 0)$. By Theorem 11.3.3 we have

$$1 = \mu_{CZ}(\Psi_{S_\infty}) = 2\alpha(S_\infty) + p(S_\infty),$$

where the parity satisfies $p(S_\infty) \in \{0, 1\}$. Therefore

$$0 = \alpha(S_\infty) = \max\{w(\eta) : \eta \in \mathfrak{S}(A_{S_\infty}) \cap (-\infty, 0)\}.$$

Therefore
$$\deg\left(t \mapsto \frac{v(Re^{2\pi it})}{|v(Re^{2\pi it})|}\right) \leq 0.$$
On the other hand, since $v \in \ker L_S$ it follows from Carleman's similarity principle in Lemma 13.6.3 that all local winding numbers of the map $v\colon \mathbb{C} \to \mathbb{C}$ are positive so that
$$\deg\left(t \mapsto \frac{v(Re^{2\pi it})}{|v(Re^{2\pi it})|}\right) \geq 0$$
and equality holds if and only if $v(z) \neq 0$ for every $z \in \mathbb{C}$. We conclude that if $v \in \ker L_S$ does not vanish identically, we necessarily have
$$v(z) \neq 0, \quad \forall\, z \in \mathbb{C}.$$
Now suppose that $v_1, v_2, v_3 \in \ker L_S$. It remains to show that the set $\{v_1, v_2, v_3\}$ is linearly dependent. Pick $z \in \mathbb{C}$. Then $v_1(z), v_2(z), v_3(z) \in \mathbb{C}$ and therefore there exist $a_1, a_2, a_3 \in \mathbb{R}$ such that
$$a_1v_1(z) + a_2v_2(z) + a_3v_3(z) = 0.$$
Because $a_1v_1 + a_2v_2 + a_3v_3 \in \ker L_S$, we conclude that
$$a_1v_1 + a_2v_2 + a_3v_3 = 0.$$
This proves that $\{v_1, v_2, v_3\}$ is linearly dependent and hence $\dim\ker L_S \leq 2$. This finishes the proof of the lemma. □

Proof of Lemma 15.6.1. By Theorem 15.5.1 and Theorem 15.5.2 it suffices to show that the operator $L_S^\delta\colon H_1^\delta \to H_0^\delta$ is surjective. By Formula (15.33) we have $\mu_{CZ}(\Psi_{S^\delta_\infty}) = 1$ and hence the lemma follows from Lemma 15.6.2. □

The proof of Lemma 15.6.2 actually reveals more. Namely, if $\zeta \neq 0 \in \ker L_S^\delta$ it follows that $\zeta(z) \neq 0$ for every $z \in \mathbb{C}$. Therefore, if $\Phi\colon \mathcal{U} \to \mathcal{M}_{\text{fast}}(\gamma)$ is the local chart from Theorem 15.5.1 the assertion of the theorem implies that
$$\{(z, z') \in \mathbb{C} \times \mathbb{C} : \widetilde{u}(z) = \Phi(\zeta)(z')\} = \{(z, z) : z \in \mathbb{C}, \zeta(z) = 0\} = \emptyset.$$
Therefore the algebraic intersection number of $\widetilde{u}$ with $\Phi(\zeta)$ satisfies
$$\mathrm{int}(\widetilde{u}, \Phi(\zeta)) = 0.$$
Suppose that γ is simple. Because in view of Theorem 14.4.5 Siefring's inequality is usually an equality and the vanishing of the algebraic intersection number occurs in an open neighborhood, we obtain
$$\mathrm{sief}(\widetilde{u}, \Phi(\zeta)) = 0.$$
Because Siefring's intersection number is a homotopy invariant, we have proved the following result.

Lemma 15.6.3. *Assume that $\widetilde{u}$ is an embedded fast finite energy plane with simple non-degenerate asymptotic orbit γ. Then its Siefring self-intersection number satisfies*

$$\mathrm{sief}(\widetilde{u}, \widetilde{u}) = 0.$$

We mention that Lemma 15.6.3 was already used to prove Theorem 14.5.5. In combination with the local automatic transversality result which we obtained in Lemma 15.6.1 we are now in position to prove the following global automatic transversality statement.

Theorem 15.6.4. *Assume that (Σ, λ) is a closed three-dimensional contact manifold satisfying $\pi_2(\Sigma) = \{0\}$ whose symplectization is endowed with an SFT-like almost complex structure. Suppose furthermore that γ is a simple non-degenerate periodic Reeb orbit with the property that there exists $[\widetilde{u}] \in \mathcal{M}_{\mathrm{fast}}(\gamma)$ such that $\widetilde{u} = (u, a)$ is embedded and $\mu_{CZ}(u) \geq 3$. Then $\mathcal{M}_{\mathrm{fast}}(\gamma)$ is a two-dimensional manifold.*

Proof. Suppose that $[\widetilde{v}] \in \mathcal{M}_{\mathrm{fast}}(\gamma)$. Because $\pi_2(\Sigma) = \{0\}$, the fast finite energy plane $\widetilde{v} = (v, b)$ is homotopic to $\widetilde{u}$. In particular,

$$\mu_{CZ}(v) = \mu_{CZ}(u) \geq 3 \quad \text{and} \quad \mathrm{sief}(\widetilde{v}, \widetilde{v}) = \mathrm{sief}(\widetilde{u}, \widetilde{u}).$$

Because $\widetilde{u}$ is embedded it follows from Theorem 14.5.5 that $\widetilde{v}$ is embedded as well. In particular, we can apply the local automatic transversality result Lemma 15.6.1 to conclude that locally around $[\widetilde{v}]$ the moduli space $\mathcal{M}_{\mathrm{fast}}(\gamma)$ is a smooth two-dimensional manifold. Because $[\widetilde{v}]$ was an arbitrary point in $\mathcal{M}_{\mathrm{fast}}(\gamma)$, the theorem is proved. $\square$

15.7 The $\mathbb{R}$-quotient

Theorem 15.6.4 tells us that the moduli space $\mathcal{M}_{\mathrm{fast}}(\gamma)$ is a manifold. On $\mathcal{M}_{\mathrm{fast}}(\gamma)$ we still have the $\mathbb{R}$-action given by $r_*[(u, a)] = [(u, a + r)]$. We know from Corollary 13.5.4 in Chapter 13 that this action is free. We show in this section that the quotient $\mathcal{M}_{\mathrm{fast}}(\gamma)/\mathbb{R}$ is a manifold as well. In order to do that we recall some facts.

The following theorem due to Godement can be found for example in [4, Theorem 3.5.25].

Theorem 15.7.1. *Assume that M is a manifold and $R \subset M \times M$ is an equivalence relation. Denote by $p_1 \colon M \times M \to M$ the projection to the first factor. Suppose that the following conditions hold*

(i) *$R \subset M \times M$ is a submanifold that is closed as a subset.*

(ii) *The restriction of the projection $p_1|_R \colon R \to M$ is a submersion.*

Then the quotient space M/R is a manifold and the quotient projection $\pi \colon M \to M/R$ is a submersion.

Remark 15.7.2. All manifolds in the theorem are assumed to satisfy the Hausdorff separation axiom. If one does not require in assertion (i) that $R \subset M \times M$ is closed, then M/R is still a manifold in the weak sense that it is only locally Euclidean and second countable, and not necessarily Hausdorff. The quotient projection is still a submersion.

In the special case where G is a Lie group action on a manifold M and the equivalence relation is the orbit relation, then

$$R = \{(x, gx) : x \in M, g \in G\}.$$

If G acts freely on M, then if $R \subset M \times M$ is closed, it is a submanifold and assertion (ii) holds. Hence we obtain the following corollary.

Corollary 15.7.3. *Assume that a Lie group G acts freely on a manifold M and $\{(x, gx) : x \in M, g \in G\}$ is closed in $M \times M$. Then the quotient M/G is a manifold and the orbit projection a submersion.*

After these preparations we are in a position to prove the main result of this section.

Theorem 15.7.4. *Under the assumptions of Theorem* 15.6.4 *it holds that the quotient $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is a one-dimensional manifold.*

Proof. By Corollary 15.7.3 it suffices to show that

$$\Big\{\big([\widetilde{u}], r_*[\widetilde{u}]\big) : [\widetilde{u}] \in \mathcal{M}_{\text{fast}}(\gamma),\ r \in \mathbb{R}\Big\} \subset \mathcal{M}_{\text{fast}}(\gamma) \times \mathcal{M}_{\text{fast}}(\gamma)$$

is closed. To see that pick a sequence $\widetilde{u}_\nu = (u_\nu, a_\nu)$ of fast finite energy planes asymptotic to γ and a sequence $r_\nu \in \mathbb{R}$ with the property that there exist fast finite energy planes $\widetilde{u} = (u, a)$ and $\widetilde{v} = (v, b)$ such that

$$\big(\widetilde{u}_\nu, (r_\nu)_*\widetilde{u}_\nu\big) \to \big(\widetilde{u}, \widetilde{v}\big).$$

This implies that

$$a_\nu \to a, \quad a_\nu + r_\nu \to b.$$

Because a_ν and b_ν both attain their minimum, it follows that there exists $r \in \mathbb{R}$ such that

$$r_\nu \to r.$$

But this implies that

$$\widetilde{v} = r_*\widetilde{u}$$

and therefore $\big\{\big([\widetilde{u}], r_*[\widetilde{u}]\big) : [\widetilde{u}] \in \mathcal{M}_{\text{fast}}(\gamma),\ r \in \mathbb{R}\big\}$ is closed in $\mathcal{M}_{\text{fast}}(\gamma) \times \mathcal{M}_{\text{fast}}(\gamma)$. This finishes the proof of the theorem. $\square$

Because a one-dimensional manifold is a disjoint union of intervals and circles, see for example [177, Appendix], we obtain the following corollary

Corollary 15.7.5. *Under the assumptions of Theorem* 15.6.4 *it holds that the quotient* $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ *is a disjoint union of intervals and circles.*

Remark 15.7.6. In view of the results about open book decompositions explained below in Section 17.1 one can actually say more. Namely, the quotient $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is either a disjoint union of intervals or a circle.

Chapter 16

Compactness

16.1 Negatively punctured finite energy planes

Assume that (Σ, λ) is a closed three-dimensional contact manifold. A *punctured holomorphic plane* is a smooth map

$$\widetilde{u} = (u, a) \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$$

where $P \subset \mathbb{C}$ is a finite subset, such that $\widetilde{u}$ satisfies the nonlinear Cauchy Riemann equation (13.2) at every point $z \in \mathbb{C} \setminus P$. In particular, if $P = \emptyset$ the map $\widetilde{u}$ is just a holomorphic plane. As in the unpunctured case the energy of a punctured holomorphic plane is defined as

$$E(\widetilde{u}) := \sup_{\phi \in \Gamma} \int_{\mathbb{C}} \widetilde{u}^* d\lambda^{\phi}.$$

A puncture $p \in P$ is called *removable* if there exists an open neighborhood of p in $\mathbb{C}$ such that the restriction of a to this neighborhood is bounded. The reason for this terminology comes from the fact that if the energy of $\widetilde{u}$ is finite, then by the theorem on removal of singularities [105, 173, 138] the map $\widetilde{u}$ can be smoothly extended over a removable puncture. If a puncture $p \in P$ is not removable, then it is called a *non-removable puncture*. For later use, we state the theorem on the removal of singularities. This version here is a special case of the version found in [138], Theorem 2.1.

Theorem 16.1.1 (Removal of singularities)**.** *Consider a symplectic manifold (M, ω) with compatible almost complex structure J. Assume that $f : D \setminus \{0\} \to M$ is a J-holomorphic map from the punctured unit disk with relatively compact image. Suppose furthermore that f has finite area, so $\int_D f^*\omega$ is finite. Then f can be extended to a J-holomorphic map $\bar{f} : D \to M$. In particular, $\bar{f}$ is smooth.*

If one thinks of $S^2 = \mathbb{C} \cup \{\infty\}$, then one can interpret a punctured holomorphic plane $\widetilde{u}$ as a map

$$\widetilde{u} \colon S^2 \setminus (P \cup \{\infty\}) \to \Sigma \times \mathbb{R}.$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_16

In particular, we might think of $\widetilde{u}$ as a *punctured holomorphic sphere* with the point $\{\infty\} \in S^2$ as an additional puncture.

Assume that $\widetilde{u} = (u, a) \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ is a punctured holomorphic plane whose energy $E(\widetilde{u})$ is finite and all of whose punctures are non-removable. It is shown in [124] that for each puncture $p \in P$ there exists an open neighborhood U of p such that the restriction of a to U is either bounded from below or above. Of course, since p is assumed to be a non-removable puncture in the first case a is unbounded from above and in the second case a is unbounded from below. This fact allows us to write the set of non-removable punctures of a holomorphic plane of finite energy as a disjoint union

$$P = P_+ \cup P_-,$$

where P_+ is the subset of punctures on which a remains bounded from below in a small neighborhood and $P_- = P \backslash P_+$ is the subset of punctures on which a remains bounded from above in a small neighborhood. Elements $p \in P_+$ are called *positive punctures*, while elements $p \in P_-$ are called *negative punctures*. Interpreting $\widetilde{u}$ as a punctured holomorphic sphere the same classification applies to the point at infinity, so that $\{\infty\}$ is either a removable, positive or negative puncture.

A special instance of a punctured holomorphic plane of finite energy is an *orbit cylinder*. Namely, if γ is a periodic Reeb orbit of period τ, define $\widetilde{\gamma} \colon \mathbb{C} \backslash \{0\} \to \Sigma \times \mathbb{R}$

$$\widetilde{\gamma}(e^{2\pi(s+it)}) = (\gamma(t), \tau s), \quad (s, t) \in \mathbb{R} \times S^1.$$

We are now in position to define

Definition 16.1.2. A *negatively punctured finite energy plane* $\widetilde{u} \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ is a punctured holomorphic plane satisfying

(i) $0 < E(\widetilde{u}) < \infty$.

(ii) All punctures $p \in P$ are non-removable and negative.

(iii) If one interprets $\widetilde{u}$ as a punctured holomorphic sphere, the point at infinity becomes an additional positive puncture.

(iv) $\widetilde{u}$ is not a reparametrization of an orbit cylinder, i.e., there does not exist $(\rho, \tau) \in \mathscr{S} = \mathbb{C}^* \ltimes \mathbb{C}$ and a periodic Reeb orbit γ such that $(\rho, \tau)_* \widetilde{u} = \widetilde{\gamma}$.

Remark 16.1.3. We mention that condition (iii) in the definition of a negatively punctured finite energy plane follows from condition (ii) and the maximum principle for a holomorphic curve $\widetilde{u} = (u, a)$. The maximum principle says that the function a does not attain a local maximum. To see that it holds, we verify $\Delta a \geq 0$ using the nonlinear Cauchy–Riemann equation for $\tilde{u}$, which reads $(d\tilde{u})^{0,1} = 0$. In components,

$$a_s + \lambda(u_t) = 0, \quad a_t - \lambda(u_s) = 0, \quad \pi_\xi u_s + J_\xi \pi_\xi u_t = 0.$$

Hence

$$\begin{aligned} du^*\lambda = d\left(\lambda(u_s)ds + \lambda(u_t)dt\right) &= \left(\frac{\partial\lambda(u_t)}{\partial s} - \frac{\partial\lambda(u_t)}{\partial s}\right) ds \wedge dt \\ &= (a_{ss} + a_{tt}) ds \wedge dt. \end{aligned}$$

On the other hand, because J is a compatible complex structure for ξ, we find that $d\lambda(u_s, u_s) = d\lambda(u_s, Ju_s) \geq 0$, so the coefficient of $ds \wedge dt$ is non-negative. This shows that $\Delta a \geq 0$, which establishes the maximum principle.

16.2 Weak SFT-compactness

Assume that $\widetilde{u} = (u, a) \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ is a negatively punctured finite energy plane. Hofer's theorem (Theorem 13.2.4) can still be applied in the case where $\widetilde{u}$ is punctured and we conclude that there exists a periodic Reeb orbit γ and a sequence s_k going to infinity such that

$$\lim_{k\to\infty} u(e^{2\pi(s_k+it)}) = \gamma(t)$$

uniformly in the C^∞-topology. If γ is non-degenerate, we refer to the negatively punctured finite energy plane as a *non-degenerate negatively punctured finite energy plane*. If $\widetilde{u}$ is a non-degenerate negatively punctured finite energy plane, it still admits an asymptotic representative. This asymptotic representative U can either be chosen to decay exponentially like in (13.8) or to vanish identically. In particular, we can associate to a non-degenerate negatively punctured finite energy plane $\widetilde{u} = (u, a)$ with asymptotic periodic orbit γ an element

$$\eta_{\widetilde{u}} = \eta_u \in [-\infty, 0),$$

where either η_u is a negative eigenvalue of the asymptotic operator A_γ such that the asymptotic representative decays exponentially with weight η_u or $\eta_u = -\infty$ and the asymptotic representative vanishes identically.

Theorem 16.2.1 (Weak SFT-compactness)**.** *Assume that γ is a non-degenerate Reeb orbit and $\widetilde{u}_\nu = (u_\nu, a_\nu) \in \widehat{\mathcal{M}}(\gamma)$ for $\nu \in \mathbb{N}$ a sequence of finite energy planes with asymptotic Reeb orbit γ. Then there exists a subsequence ν_j, a sequence of gauge transformations $(r_j, (\rho_j, \tau_j)) \in \mathbb{R} \times \mathscr{S}$ and a negatively punctured finite energy plane $\widetilde{u} = (u, a) \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ with positive asymptotic orbit γ such that $(r_j, (\rho_j, \tau_j))_* \widetilde{u}_{\nu_j}$ converges in the C^∞_{loc}-topology to $\widetilde{u}$. Moreover,*

$$\eta_u \leq \eta_{u_{\nu_j}}, \quad j \in \mathbb{N}. \tag{16.1}$$

This result is a special case of the SFT-compactness theorem, see [43, 121, 125]. We make the following remarks.

Remark 16.2.2. The fact that $\widetilde{u}_\nu \in \widehat{\mathcal{M}}(\gamma)$ implies that

$$E(\widetilde{u}_\nu) = \tau$$

where τ is the period of the periodic orbit γ. In particular, the energy of the sequence $\widetilde{u}_\nu$ is constant and therefore uniformly bounded.

Remark 16.2.3. That in the limit no positive punctures occur follows from the maximum principle mentioned in Remark 16.1.3.

Observe, that in view of Theorem 16.2.1 in order to show that the moduli space $\widehat{\mathcal{M}}(\gamma)/\mathbb{R} \times \mathscr{S} = \mathcal{M}(\gamma)/\mathbb{R}$ is compact it suffices to show that the limit has no negative punctures, i.e., is an honest finite energy plane.

16.3 The systole

Assume that (Σ, λ) is a closed contact manifold. Denote by

$$\mathcal{R} = \mathcal{R}(\Sigma, \lambda) \subset C^\infty(S^1, \Sigma)$$

the set of all periodic Reeb orbits of Σ, i.e., the set of all loops $\gamma \in C^\infty(S^1, \Sigma)$ for which there exists a positive number $\tau = \tau_\gamma$, referred to as the period, such that the pair (γ, τ) is a solution of the ODE $\partial_t \gamma = \tau R(\gamma)$. The *systole* of (Σ, λ) is defined as

$$\mathrm{sys}(\Sigma, \lambda) = \inf\{\tau_\gamma : \gamma \in \mathcal{R}\}.$$

Here we use the convention that infimum of the empty set equals infinity. However, in view of Weinstein's conjecture the systole of every closed contact manifold is expected to be finite. In view of the result by Taubes [233] this is definitely true in dimension three. Moreover, because our contact manifold is assumed to be closed, it follows from the theorem of Arzelà–Ascoli that if a periodic Reeb orbit exists the infimum is actually attained so that for a three-dimensional contact manifold we can define the systole also as a minimum

$$\mathrm{sys}(\Sigma, \lambda) = \min\{\tau_\gamma : \gamma \in \mathcal{R}\}.$$

We further point out that if the contact manifold (N, λ) satisfies in addition

$$H_1(\Sigma; \mathbb{Q}) = \{0\},$$

then the systole only depends on the Hamiltonian structure $(\Sigma, d\lambda)$ and not on the choice of the contact form λ. Indeed, if λ and λ' are two contact forms on Σ satisfying

$$d\lambda = d\lambda',$$

then since the first rational homology group of Σ vanishes there exists a smooth function $f \in C^\infty(\Sigma, \mathbb{R})$ such that

$$\lambda = \lambda' + df.$$

Now in view of Stokes' theorem and the definition of the Reeb vector field we compute for $\gamma \in \mathcal{R}$

$$\tau_\gamma = \int_{S^1} \gamma^* \lambda = \int_{S^1} \gamma^* \lambda'.$$

This proves that if the first rational homology group vanishes, the systole only depends on the Hamiltonian structure $(\Sigma, d\lambda)$.

If $\gamma \in \mathcal{R}$ satisfies $\tau_\gamma = \mathrm{sys}(\Sigma, \lambda)$, we say that the periodic Reeb orbit γ *represents the systole* of (Σ, λ). Note that the systole in general does not have a unique representative. However, each representative of the systole necessarily has minimal period among all periodic Reeb orbits.

Theorem 16.3.1. *Assume that (Σ, λ) is a closed, three-dimensional contact manifold and γ is a non-degenerate Reeb orbit of (Σ, λ) which represents the systole. Then the moduli space $\mathcal{M}(\gamma)/\mathbb{R}$ is compact.*

Proof. In view of Theorem 16.2.1 it suffices to show that each negatively punctured finite energy plane $\widetilde{u} = (u, a)\colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ has no punctures, i.e., $P = \emptyset$. We argue by contradiction and assume that $P \neq \emptyset$. Hence suppose

$$P = \{p_1, \dots, p_\ell\}$$

for $\ell \in \mathbb{N}$. Hofer's theorem (Theorem 13.2.4) can also be applied to negative punctures. For a negative puncture at p_j with $j \in \{1, \dots, \ell\}$ it asserts that there exists a sequence s_k^j going to $-\infty$ and a periodic Reeb orbit γ_j such that

$$\lim_{k \to \infty} u(e^{2\pi(s_k^j + it)} + p_j) = \gamma_j(t) \tag{16.2}$$

uniformly in the C^∞-topology. Because the puncture at infinity is positive, there exists moreover a sequence s_k going to infinity such that

$$\lim_{k \to \infty} u(e^{2\pi(s_k + it)}) = \gamma(t) \tag{16.3}$$

uniformly in the C^∞-topology. Because $\widetilde{u}$ is holomorphic, the following holds by a special instance of Formula (13.4)

$$u^* d\lambda = |\pi \partial_x u|^2 \geq 0. \tag{16.4}$$

Abbreviate by $D_R(p) = \{z \in \mathbb{C} : |z - p| \leq R\}$ the disk of radius R centered at p. There exists $k_0 \in \mathbb{N}$ such that for every $k \geq k_0$ and every $1 \leq j, j' \leq \ell$ satisfying $j \neq j'$ we have

$$D_{e^{2\pi s_k^j}}(p_j) \cap D_{e^{2\pi s_k^{j'}}}(p_j') = \emptyset, \quad D_{e^{2\pi s_k^j}}(p_j) \subset D_{e^{2\pi s_k}}(0).$$

In view of Stokes' theorem we conclude for every $k \geq k_0$

$$0 \leq \int_{D_{e^{2\pi s_k}}(0) \setminus \bigcup_{j=1}^{\ell} D_{e^{2\pi s_k^j}}(p_j)} u^* d\lambda = \int_{\partial D_{e^{2\pi s_k}}(0)} u^*\lambda - \sum_{j=1}^{\ell} \int_{\partial D_{e^{2\pi s_k^j}}(p_j)} u^*\lambda.$$

Since this is true for any $k \geq k_0$, we obtain in view of (16.2) and (16.3)

$$0 \leq \int_{S^1} \gamma^*\lambda - \sum_{j=1}^{\ell} \int_{S^1} \gamma_j^*\lambda = \tau_\gamma - \sum_{j=1}^{\ell} \tau_{\gamma_j} \tag{16.5}$$

where τ_{γ_j} are the periods of the periodic orbits γ_j. Because γ represents the systole, the following inequalities hold true

$$\tau_{\gamma_j} \geq \tau_\gamma, \quad 1 \leq j \leq \ell. \tag{16.6}$$

From (16.5) and (16.6) we conclude

$$\ell = 1, \quad \tau_{\gamma_1} = \tau_\gamma.$$

In view of (16.4) we further get

$$\pi \partial_x u(z) = 0, \quad z \in \mathbb{C}$$

and because $\widetilde{u}$ is holomorphic, we also get

$$\pi \partial_y u(z) = 0 \quad z \in \mathbb{C}.$$

In particular, $\widetilde{u}$ is an orbit cylinder up to reparametrization, in contradiction to assertion (iv) in Definition 16.1.2. This finishes the proof of the theorem. □

Corollary 16.3.2. *Suppose the assumptions of Theorem* 16.3.1. *Then the moduli space* $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ *is compact.*

Proof. Recall from Corollary 13.6.8 that a non-degenerate finite energy plane $\widetilde{u} = (u, a)$ is fast if and only if the winding number of its asymptotic eigenvalue satisfies $w(\eta_u) = 1$. By Theorem 13.6.1 the inequality $w(\eta_u) \geq 1$ holds. If $\widetilde{u}$ is the limit of fast finite energy planes, it follows from (16.1) that $w(\eta_u) \leq 1$. Therefore, it follows that $w(\eta_u) = 1$ and $\widetilde{u}$ is fast. We have shown that the moduli space $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is closed in $\mathcal{M}(\gamma)/\mathbb{R}$. Now the corollary follows from Theorem 16.3.1. □

In the following theorem we weaken the assumption that the binding orbit γ has minimal period.

Theorem 16.3.3. *Assume that* $\Sigma \subset \mathbb{C}^2$ *is a starshaped hypersurface and* γ *is a non-degenerate simple Reeb orbit on* Σ *with the property that* γ *is linked to every Reeb orbit on* Σ *of smaller period than* γ. *Suppose furthermore that there exists a fast finite energy plane of Siefring self-intersection number* 0 *asymptotic to* γ *and therefore all fast finite energy planes asymptotic to* γ *have vanishing Siefring self-intersection number. Then the moduli space* $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ *is compact.*

Proof. As in the proof of Theorem 16.3.1 we argue by contradiction and suppose that there exists a negatively punctured finite energy plane $\widetilde{u} = (u, a) \colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ with nonempty set of negative punctures $P = \{p_1, \ldots, p_\ell\}$ which arises as the weak SFT limit of a sequence of fast finite energy planes asymptotic to γ. For $j \in \{1, \ldots, \ell\}$ choose periodic Reeb orbits γ_j such that (16.2) holds true. As the proof of Theorem 16.3.1 reveals, it holds that

$$\tau_{\gamma_j} < \tau_\gamma. \tag{16.7}$$

In particular, γ_j is disjoint from γ. Hence we can choose an open neighborhood $U_j \subset \Sigma$ of γ_j which contracts to γ_j and is disjoint from γ. Because $\widetilde{u}$ arises as the weak SFT-limit of fast finite energy planes, there exists in view of (16.2) $\widetilde{v} = (v, b) \in \widehat{\mathcal{M}}_{\mathrm{fast}}(\gamma)$ such that

$$\mathrm{im}(v|_{\partial D}) \subset U_j, \quad v|_{\partial D} \sim_{U_j} \gamma_j \tag{16.8}$$

where $D = \{z \in \mathbb{C} : ||z|| \leq 1\}$ is the unit sphere in $\mathbb{C}$ and $\sim_{U_j}$ means homotopic in U_j. By assumption of the theorem it holds that

$$\mathrm{sief}(\widetilde{v}, \widetilde{v}) = 0.$$

Therefore in view of Theorem 14.5.5 in Chapter 14 we conclude that

$$\mathrm{im}(v) \cap \mathrm{im}(\gamma) = \emptyset.$$

Combining this fact with (16.8) and the choice of the neighborhood U_j, we conclude that there exists $w \in C^\infty(D, \Sigma)$ with the property that

$$w(e^{2\pi i t}) = \gamma_j(t),\ t \in S^1, \quad \mathrm{im}(w) \cap \mathrm{im}(\gamma) = \emptyset.$$

This shows that γ and γ_j are not linked. By assumption of the theorem this implies that $\tau_{\gamma_j} \geq \tau_\gamma$. However, this contradicts (16.7) and this contradiction proves the theorem. $\square$

16.4 Dynamical convexity

Recall that a periodic Reeb orbit $\gamma \in C^\infty(S^1, \Sigma)$ of period τ is called non-degenerate if $\det(d^\xi \phi_R^\tau(\gamma(0)) - \mathrm{id}) \neq 0$.

Definition 16.4.1. A contact manifold (Σ, λ) is called *non-degenerate* if all periodic Reeb orbits on (Σ, λ) are non-degenerate.

After a small perturbation we can always assume that a closed contact manifold is non-degenerate. To make this statement precise, recall that if $f \in C^\infty(\Sigma, \mathbb{R}_+)$ is a smooth positive function on Σ, then the one-form $\lambda_f := f\lambda \in \Omega^1(\Sigma)$ is still a contact form on Σ. Note that the contact structure $\xi = \ker \lambda = \ker \lambda_f$ remains unchanged under this procedure, although the Reeb vector field and therefore the dynamics on Σ might change dramatically. The following result is due to Robinson [213], see also [126, Proposition 6.1].

Theorem 16.4.2. *Assume that (Σ, λ) is a closed contact manifold. Then there exists a subset $\mathcal{F} \subset C^\infty(\Sigma, \mathbb{R}_+)$ which can be written as a countable intersection of open and dense subsets of $C^\infty(\Sigma, \mathbb{R}_+)$ such that for every $f \in \mathcal{F}$ the contact form λ_f is non-degenerate.*

Since $\mathcal{F}$ is a countable intersection of open and dense subsets, it follows from Baire's theorem that $\mathcal{F}$ is dense itself. This explains why after a small perturbation we can assume that the contact manifold is non-degenerate.

Suppose that γ is a contractible closed Reeb orbit in a contact manifold (Σ, λ) of period τ. Since γ is contractible, there exists a *filling disk* for γ, i.e., a smooth map $\overline{\gamma}\colon D = \{z \in \mathbb{C} : |z| \leq 1\} \to \Sigma$ such that

$$\overline{\gamma}(e^{2\pi i t}) = \gamma(t), \quad t \in S^1.$$

Choose a symplectic trivialization

$$\mathfrak{T}\colon \overline{\gamma}^* \xi \to D \times \mathbb{C}.$$

We define the Conley–Zehnder index of the filling disk $\overline{\gamma}$ as

$$\mu_{CZ}(\overline{\gamma}) = \mu_{CZ}\Big(t \mapsto \mathfrak{T}_{e^{2\pi i t}} d^\xi \phi_R^{t\tau}(\gamma(0))\mathfrak{T}_1^{-1}\Big).$$

The Conley–Zehnder index is independent of the choice of the symplectic trivialization and depends only on the homotopy class of the filling disk $\overline{\gamma}$. If $\overline{\gamma}'$ is another filling disk for γ, one obtains a sphere $\overline{\gamma}\#(\overline{\gamma}')^-$ by gluing $\overline{\gamma}$ and $(\overline{\gamma}')^-$, the filling disk $\overline{\gamma}'$ with opposite orientation, along γ. In view of Lemma 15.2.2 the Conley–Zehnder indices with respect to the two filling disks are related by

$$\mu_{CZ}(\overline{\gamma}) - \mu_{CZ}(\overline{\gamma}') = 2c_1\Big((\overline{\gamma}\#(\overline{\gamma}')^-)^*\xi\Big). \tag{16.9}$$

The first Chern number gives rise to a homomorphism

$$I_{c_1}\colon \pi_2(\Sigma) \to \mathbb{Z}, \quad [v] \mapsto c_1(v^*\xi).$$

Suppose now that the homomorphism I_{c_1} is trivial. This for example happens if $\pi_2(\Sigma) = \{0\}$. Then it follows from (16.9) that the Conley–Zehnder index is independent of the choice of the filling disk $\overline{\gamma}$ and only depends on the periodic Reeb orbit γ. Hence under the assumption that $I_{c_1} = 0$ we can set

$$\mu_{CZ}(\gamma) := \mu_{CZ}(\overline{\gamma})$$

where $\overline{\gamma}$ is any filling disk for the contractible Reeb orbit γ.

Definition 16.4.3. A closed three-dimensional contact manifold (Σ, λ) is called *dynamically convex* if $I_{c_1} = 0$ and every closed contractible Reeb orbit γ of Σ satisfies

$$\mu_{CZ}(\gamma) \geq 3.$$

Remark 16.4.4. It was shown by Hofer, Wysocki, and Zehnder in [127], that the second homotopy group of a dynamically convex manifold is trivial, so that instead of assuming that $I_{c_1} = 0$ vanishes one can equivalently assume that $\pi_2 = \{0\}$.

If γ is a periodic Reeb orbit recall that the *covering number* of γ is defined as

$$\mathrm{cov}(\gamma) = \max\Big\{k \in \mathbb{N} : \gamma(t + \tfrac{1}{k}) = \gamma(t),\ \forall\, t \in S^1\Big\}.$$

Moreover, a periodic Reeb orbit γ is called *simple* if $\mathrm{cov}(\gamma) = 1$. The following theorem is due to Hryniewicz [133].

Theorem 16.4.5 (Hryniewicz)**.** *Assume that (Σ, λ) is a non-degenerate, dynamically convex closed three-dimensional contact manifold and γ is a simple periodic Reeb orbit of Σ. Then the moduli space $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is compact.*

Proof. Suppose that $\widetilde{u}_\nu = (u_\nu, a_\nu)$ is a sequence of fast finite energy planes which asymptotic orbit γ which converge to a negatively punctured finite energy plane $\widetilde{u} = (u, a)\colon \mathbb{C} \setminus P \to \Sigma \times \mathbb{R}$ with asymptotic orbit γ in the C^∞_{loc}-topology. It remains to show that the set of negative punctures P is empty and that $\widetilde{u}$ is fast. We first rule out the danger that $\widetilde{u}$ is a so-called connector, namely a negatively punctured finite energy plane satisfying

$$||\pi\partial_x u||^2 = \frac{1}{2}\Big(||\pi\partial_x u||^2 + ||\pi\partial_y u||^2\Big) = 0$$

where $\pi\colon T\Sigma \to \xi$ is the projection along the Reeb vector field. If $\widetilde{u}$ is a connector, it follows that

$$\widetilde{u} = \widetilde{\gamma'} \circ p,$$

where $\widetilde{\gamma'}$ is the orbit cylinder over a periodic Reeb orbit γ' and $p\colon \mathbb{C} \to \mathbb{C}$ is a holomorphic map, which has to be a polynomial because the energy of $\widetilde{u}$ is finite. Because γ is simple, it follows that $\gamma' = \gamma$ and p has degree one, i.e., $\widetilde{u}$ is a reparametrization of an orbit cylinder which is forbidden by condition (iv) in Definition 16.1.2. Therefore $\widetilde{u}$ is not a connector.

We now suppose by contradiction that the set of punctures is not empty, so that we can write $P = \{p_1, \dots, p_\ell\}$ for $\ell \in \mathbb{N}$. Because (Σ, λ) is non-degenerate and $\widetilde{u}$ is not a connector, the asymptotic description from Theorem 13.5.2 can now be applied to the negative punctures as well. Namely for each $1 \leq j \leq \ell$ there exists a periodic Reeb orbit γ_j, a positive eigenvalue η_j of the operator A_{γ_j} and an eigenvector ζ_j of A_{γ_j} to the eigenvalue η_j such that the puncture p_j admits an asymptotic representative of the form

$$U_j(s,t) = e^{\eta_j s}\big(\zeta_j(t) + \kappa_j(s,t)\big)$$

where each function κ_j decays with all derivatives exponentially with uniform exponential weight. For negative punctures asymptotic representative means that there exist proper embeddings $\phi_j \colon (-\infty, R_j] \times S^1 \to \mathbb{R} \times S^1$ asymptotic to the identity such that

$$\widetilde{u}\Big(e^{\phi_j(s,t)} + p_j\Big) = \Big(\exp_{\gamma_j(t)} U_j(s,t), \tau_j s\Big),$$

where exp is the exponential map for some Riemannian metric on Σ and τ_j are the periods of the periodic orbits γ_j. Because $\widetilde{u}$ is the limit of the finite energy planes $\widetilde{u}_\nu$, it follows that the periodic orbits γ_j are contractible. Hence we can pick for each periodic orbit γ_j a filling disk $\overline{\gamma_j}$. Pick unitary trivializations $\mathfrak{T}_j \colon \overline{\gamma_j}^* \xi \to D \times \mathbb{C}$, i.e., trivializations which are complex with respect to the complex structure J on ξ and orthogonal with respect to the metric $\omega(\cdot, J\cdot)$ on ξ. The restriction of the trivializations to the periodic Reeb orbits γ_j gives rise to the bounded linear operators

$$A_{\gamma_j}^{\mathfrak{T}_j} \colon W^{1,2}(S^1, \mathbb{C}) \to L^2(S^1, \mathbb{C})$$

as in (13.7). Because the operator $A_{\gamma_j}^{\mathfrak{T}_j}$ is conjugated to the operator A_{γ_j}, the eigenvalue η_j of the operator A_{γ_j} can also be interpreted as an eigenvalue of the operator $A_{\gamma_j}^{\mathfrak{T}_j}$. As an eigenvalue of the operator $A_{\gamma_j}^{\mathfrak{T}_j}$ it has a winding number

$$w(\eta_j, \overline{\gamma}_j) \in \mathbb{Z}.$$

As the notation indicates, the winding number is independent of the choice of the trivialization $\mathfrak{T}_j$. A priori it depends at least up to homotopy on the choice of the filling disk $\overline{\gamma_j}$. If $\overline{\gamma_j}'$ is another filling disk, then the winding numbers are related by

$$w(\eta_j, \overline{\gamma_j}) - w(\eta_j, \overline{\gamma_j}') = c_1\Big((\overline{\gamma_j} \# (\overline{\gamma_j}')^-)^* \xi\Big).$$

Observe now that the homomorphism $I_{c_1} \colon \pi_2(\Sigma) \to \mathbb{Z}$ is trivial, since this is a necessary requirement to even define the dynamical convexity of (Σ, λ) in Definition 16.4.3. Hence the winding number is independent of the choice of the filling disk, so we can set

$$w(\eta_j) := w(\eta_j, \overline{\gamma_j}).$$

Using again that (Σ, λ) is dynamically convex it holds that

$$\mu_{CZ}(\gamma_j) \geq 3, \quad 1 \leq j \leq \ell.$$

Because η_j is positive, we conclude in view of Theorem 11.3.3 and the monotonicity of the winding number from Corollary 11.3.2 that

$$w(\eta_j) \geq 2, \quad 1 \leq j \leq \ell. \tag{16.10}$$

By gluing the filling disks $\overline{\gamma_j}$ to u along γ_j for $1 \leq j \leq \ell$ we obtain an open disk

$$u \# \bigcup_{j=1}^{\ell} \overline{\gamma_j}$$

whose closure is a filling disk for γ. Choose a trivialization

$$\mathfrak{T}\colon u^*\xi \to (\mathbb{C} \setminus P) \times \mathbb{C}$$

which extends at the positive puncture to a trivialization $\mathfrak{T}\colon \gamma^*\xi \to S^1 \times \mathbb{C}$ and coincides at the negative punctures with $\mathfrak{T}_j\colon \gamma_j^*\xi \to S^1 \times \mathbb{C}$. Inspired by the proof of Theorem 13.6.1 we consider the smooth map

$$\mathfrak{T}\pi\partial_x u\colon \mathbb{C} \setminus P \to \mathbb{C}.$$

In view of (16.10) there exists $\epsilon > 0$ such that for the loops

$$\gamma_j^\epsilon\colon S^1 \to \mathbb{C}, \quad t \mapsto p_j + \epsilon e^{2\pi i t}$$

where $1 \leq j \leq \ell$ the winding number as defined in (13.11) of the map $\mathfrak{T}\pi\partial_r u$ along these loops satisfies

$$w_{\gamma_j^\epsilon}(\mathfrak{T}\pi\partial_r u) \geq 2.$$

In view of (13.18) we conclude that

$$w_{\gamma_j^\epsilon}(\mathfrak{T}\pi\partial_x u) \geq 1.$$

Because $\widetilde{u}$ is not a connector, it follows that η_u is finite and therefore an eigenvalue of the asymptotic operator A_γ. Because $\widetilde{u}$ is the limit of fast finite energy planes, it follows from (16.1) and the monotonicity of winding numbers established in Corollary 11.3.2 that

$$w(\eta_u) \leq 1.$$

Hence there exists $R > 0$ such that

$$w_R(\mathfrak{T}\pi\partial_r u) \leq 1,$$

where we recall from (13.12) that w_R denotes the winding number of the loop $t \mapsto Re^{2\pi i t}$. Again using (13.18) we conclude that

$$w_R(\mathfrak{T}\pi\partial_x u) \leq 0.$$

On the other hand, since $\widetilde{u}$ is holomorphic, we conclude using Carleman's similarity principle as in the proof of Theorem 13.6.1 that

$$0 \geq w_R(\mathfrak{T}\pi\partial_x u) \geq \sum_{j=1}^{\ell} w_{\gamma_j^\epsilon}(\mathfrak{T}\pi\partial_x u) \geq \ell.$$

This implies that $\ell = 0$. Hence $\widetilde{u}$ has no negative punctures and is therefore a finite energy plane. In view of Theorem 13.6.5 the winding number of its asymptotic eigenvalue satisfies $w(\eta_u) \geq 1$ and because $\widetilde{u}$ is the limit of fast finite energy planes, it follows from (16.1) that $w(\eta_u) = 1$. This shows that $\widetilde{u}$ is fast and the theorem is proved. $\square$

Chapter 17

Construction of Global Surfaces of Section

17.1 Open book decompositions

Theorem 17.1.1. *Assume that $\Sigma \subset \mathbb{C}^2$ is a starshaped hypersurface and $\gamma \in C^\infty(S^1, \Sigma)$ is a non-degenerate periodic Reeb orbit of period τ satisfying $\mu_{CZ}(\gamma) \geq 3$. Suppose furthermore that γ is linked to every periodic Reeb orbit on Σ of period less than τ. If under these assumptions $\widetilde{u} = (u, a)$ is a fast finite energy plane with asymptotic orbit γ and vanishing Siefring self-intersection number, then $u \colon \mathbb{C} \to \Sigma$ is a global surface of section.*

Proof. The basic ingredients of the proof of this theorem appeared first in [126]. Because $\widetilde{u}$ is fast the image of u is transverse to the Reeb vector field. Moreover, because its Siefring self-intersection number vanishes, it follows from Theorem 14.5.5 that

$$\mathrm{im}(u) \cap \mathrm{im}(\gamma) = \emptyset.$$

Therefore it suffices to check the global condition of a global surface of section, namely condition (iii) of Definition 9.1.1. Consider the quotient of the moduli space of fast finite energy planes $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$. By Theorem 15.7.4 this is a one-dimensional manifold. Moreover, it follows from Theorem 16.3.3 that it is compact as well. We examine the subset

$$\mathcal{I} := \bigcup_{\widetilde{v}=(v,b)\in\widehat{\mathcal{M}}_{\text{fast}}(\gamma)} \mathrm{im}(v) \subset \Sigma.$$

Because by assumption the Siefring self-intersection number $\widetilde{u} = (u, a)$ satisfies $\mathrm{sief}(\widetilde{u}, \widetilde{u}) = 0$, the same is true for every other fast finite energy plane $\widetilde{v} = (v, b)$ asymptotic to γ. Therefore it follows from Theorem 14.5.5 that $\mathrm{im}(v) \cap \mathrm{im}(\gamma) = \emptyset$ and consequently

$$\mathcal{I} \subset \Sigma \setminus \mathrm{im}(\gamma).$$

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_17

Again from Theorem 14.5.5 it follows that $v\colon \mathbb{C} \to \Sigma$ is an embedding. Moreover, from Corollary 14.5.7 we know that the images of u and v are disjoint unless they coincide. In particular, it follows that $\mathcal{I}$ is open in $\Sigma \setminus \mathrm{im}(\gamma)$.

For more detail, take $p \in \mathcal{I}$, and take $\tilde{v}_0 = (v_0, b_0) \in \widehat{\mathcal{M}}_{\text{fast}}(\gamma)$ and $z \in \mathbb{C}$ such that $v_0(z) = p$. We get a point $[\tilde{v}_0]$ in the one-dimensional moduli space $\mathcal{M}_{\text{fast}}(\gamma)$. Take a smooth, regular and injective curve $s \mapsto [\tilde{v}_s]$ through this point defined on a short interval $(-\epsilon, \epsilon)$. By Corollary 15.7.3 and Theorem 15.5.1 we can lift this curve to a curve $s \mapsto \tilde{v}_s = (v_s, b_s)$ in $\widehat{\mathcal{M}}_{\text{fast}}(\gamma)$. This allows us to consider a local version of the evaluation map by putting

$$ev : (-\epsilon, \epsilon) \times \mathbb{C} \to \Sigma \setminus \gamma, \quad (s, w) \mapsto \tilde{v}_s(z).$$

This map is continuous and maps an open set in $\mathbb{R}^3$ to a 3-manifold. Furthermore, the map ev is injective because of the previously mentioned Corollary 14.5.7. By invariance of domain, ev is open. Explicitly $ev((-\epsilon, \epsilon), B_\epsilon(z))$ is an open neighborhood of p in $\mathcal{I}$.

Because $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is compact, it is closed as well. By assumption it is nonempty and since $\Sigma \setminus \mathrm{im}(\gamma)$ is connected, it follows that

$$\mathcal{I} = \Sigma \setminus \mathrm{im}(\gamma).$$

This means that for every point $x \in \Sigma \setminus \mathrm{im}(\gamma)$ there exists a unique $[\widetilde{v}] = [(v, b)] \in \mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ such that $x \in \mathrm{im}(v)$. This gives rise to a projection

$$\pi\colon \Sigma \setminus \mathrm{im}(\gamma) \to \mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}. \tag{17.1}$$

In particular, since $\Sigma \setminus \mathrm{im}(\gamma)$ is connected we conclude that the one-dimensional manifold $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ satisfies

$$\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R} \cong S^1.$$

This construction endows Σ with the structure of a so-called *planar open book* .

Definition 17.1.2. By this we mean that Σ comes with a codimension 2-submanifold B, called *binding*, which has a neighborhood that is trivialized by $\epsilon : B \times D^2 \to \Sigma$, and fiber bundle $\theta : \Sigma \backslash B \to S^1$, which is standard with respect to the trivialization ϵ, meaning

$$\theta \circ \epsilon : B \times D^2 \setminus \{0\} \longmapsto S^1, \quad (b; r, \phi) \longmapsto \phi.$$

We call the closures of these fibers *pages*. The adjective *planar* means that the fibers of θ are surfaces that can be embedded in the plane.

In this case, the pages are disks, which are the closures of the images of the maps v. The binding is the periodic Reeb orbit γ. Moreover, the Reeb vector field is transverse to the pages. It remains to show that the flow of the Reeb vector field maps each page of the open book back to itself in future and past. We

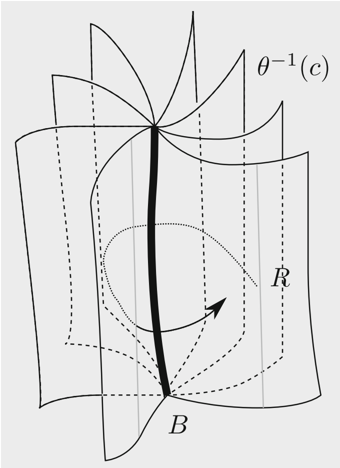


Figure 17.1: An open book with a Reeb flow transverse to the pages.

concentrate our argument to the future; the past is completely analogous. Hence we pick a point $x \in \mathrm{im}(u)$ and we have to show that there exists $t > 0$ such that $\phi_R^t(x) \in \mathrm{im}(u)$ again. We consider the Omega-limit set of x defined as follows

$$\Omega(x) := \bigcap_{\tau=0}^{\infty} \overline{\{\phi_R^t(x) : t \geq \tau\}} \subset \Sigma.$$

We first consider the case

$$\Omega(x) \cap \mathrm{im}(\gamma) = \emptyset. \tag{17.2}$$

In this case there exists a compact subset $K \subset \Sigma \setminus \mathrm{im}(\gamma)$ such that

$$\phi_R^t(x) \in K, \quad t \geq 0.$$

Because R is transverse to every page of the open book decomposition (17.1) we can identify $\mathcal{M}_{\mathrm{fast}}(\gamma)/\mathbb{R}$ in such a way with the circle $S^1 = \mathbb{R}/\mathbb{Z}$ that

$$d\pi(x)R(x) > 0, \quad \forall\, x \in \Sigma \setminus \mathrm{im}(\gamma).$$

Because $K \subset \Sigma \setminus \mathrm{im}(\gamma)$ is compact, there exists $\epsilon > 0$ such that

$$d\pi(x)R(x) > \epsilon, \quad \forall\, x \in K.$$

This means that there exists $t \in \left(0, \frac{1}{\epsilon}\right)$ such that

$$\phi_R^t(x) \in \mathrm{im}(u).$$

This proves condition (iii) of a global surface of section in case (17.2) holds true. It remains to discuss the case

$$\Omega(x) \cap \mathrm{im}(\gamma) \neq \emptyset. \tag{17.3}$$

In this case, Corollary 10.6.3 together with the assumption that $\mu_{CZ}(\gamma) \geq 3$ implies that the rotation number of the linearized flow of ϕ_R^t at γ is bigger one. On the other hand because $\widetilde{u}$ is fast its asymptotic eigenvalue has winding number one and therefore it follows with the help of the asymptotic behavior of a fast finite energy plane that $\phi_R^t(x)$ has to intersect $\mathrm{im}(u)$ in forward time in the case (17.3) as well. This finishes the proof of the theorem. □

Theorem 17.1.3. *Assume that $\Sigma_r \subset \mathbb{C}^2$ for $r \in [0,1]$ is a smooth family of star-shaped hypersurfaces and $\gamma_r \in C^\infty(S^1, \Sigma_r)$ is a smooth family of non-degenerate periodic Reeb orbits. Suppose furthermore that γ_r is linked to every periodic Reeb orbit on Σ_r of period less than the one of γ_r for every $r \in [0,1]$. Moreover, assume J_r is a smooth family of SFT-like complex structures on $\Sigma_r \times \mathbb{R}$ for every $r \in [0,1]$. Now if there exists a fast finite energy plane of vanishing Siefring self-intersection number asymptotic to γ_0 with respect to the SFT-like complex structure J_0, then there also exists a fast finite energy plane of vanishing Siefring self-intersection number asymptotic to γ_1 with respect to J_1.*

Proof. Consider the moduli space

$$\mathcal{N} = \big\{(r, [\widehat{u}]) : [\widehat{u}] \in \mathcal{M}_{\mathrm{fast}}(\gamma_r; J_r)/\mathbb{R}\big\}.$$

If $\mu_{CZ}(\gamma_0) = 2$ and hence $\mu_{CZ}(\gamma_r) = 2$ for every $r \in [0,1]$, then the moduli space $\mathcal{N}$ is a one-dimensional manifold with boundary and if $\mu_{CZ}(\gamma_0) \geq 3$, then $\mathcal{N}$ is a two-dimensional manifold with boundary. The boundary of $\mathcal{N}$ is given by

$$\partial\mathcal{N} = \Big(\{0\} \times \mathcal{M}_{\mathrm{fast}}(\gamma_0, J_0)/\mathbb{R}\Big) \cup \Big(\{1\} \times \mathcal{M}_{\mathrm{fast}}(\gamma_1, J_1)/\mathbb{R}\Big).$$

By Theorem 15.7.4 the moduli space $\mathcal{N}$ is compact. By applying the Ehresmann fibration theorem, see Theorem 5.7.6, to the map

$$p\colon \mathcal{N} \to [0,1], \quad (r, [\widetilde{u}]), \mapsto r$$

we deduce that we have a diffeomorphism

$$\mathcal{M}_{\mathrm{fast}}(\gamma_0, J_0)/\mathbb{R} \cong \mathcal{M}_{\mathrm{fast}}(\gamma_1, J_1)/\mathbb{R}.$$

Because $\mathcal{M}_{\mathrm{fast}}(\gamma_0, J_0)$ is nonempty by assumption, the same holds for $\mathcal{M}_{\mathrm{fast}}(\gamma_1, J_1)$. This finishes the proof of the theorem. □

As a corollary of this result we are now in a position to prove Theorem 9.4.10 which we state for the reader's convenience as a corollary.

Corollary 17.1.4. *Assume that a pair (μ, c), consisting of a mass ratio μ and an energy value c, lies in the Birkhoff set $\mathfrak{B}$. Then the retrograde periodic orbit γ_R bounds a global surface of section on $\Sigma_{\mu,c}$.*

Proof. By definition of the Birkhoff set we can find a homotopy as in Theorem 17.1.3 to the round sphere $S^3 \subset \mathbb{C}^2$. On the round sphere explicit examples of fast finite energy planes with vanishing Siefring self-intersection number can be written down as was done in Section 9.4.1. The examples from Section 9.4.1 do not satisfy the non-degeneracy condition. They can still be used because they do converge exponentially to the binding orbit, but this requires additional work. We will not take this route though, so for the sake of completeness we will work out examples that satisfy the required non-degeneracy assumption after finishing this argument.

Assuming the existence of these examples for the moment, we complete the argument by applying Theorem 17.1.3; we deduce that there exists a fast finite energy plane of vanishing Siefring self-intersection number with asymptotic orbit the retrograde periodic orbit on $\Sigma_{\mu,c}$. The corollary follows now from Theorem 17.1.1. □

17.1.1 More examples of finite energy planes

We will now work out the required non-degenerate example needed for the completion of Corollary 17.1.4. We fix two real numbers a_1 and a_2. Later we will specialize to the case $a_1 = 1$ and a_2 is irrational. We consider the "irrational ellipsoid", meaning the Hamiltonian on $(\mathbb{R}^4, \omega = dx \wedge dy)$ given by

$$H = \sum_j \frac{a_j}{2}(x_j^2 + y_j^2).$$

Define the level set $\Sigma := H^{-1}(\frac{1}{2})$. As usual, we define the Hamiltonian vector field X_H; it is given by

$$X_H = \sum_j a_j \left(x_j \frac{\partial}{\partial y_j} - y_j \frac{\partial}{\partial x_j} \right).$$

We will also need the standard Liouville vector field, given by

$$X = \frac{1}{2} \sum_j \left(x_j \frac{\partial}{\partial x_j} + y_j \frac{\partial}{\partial y_j} \right).$$

We define two more vector fields, inspired by quaternionic multiplication with j and k, by putting $\tilde{U} = 2jX$ and $\tilde{V} = 2kX$. In components,

$$\begin{aligned}
\tilde{U} &= -x_2 \frac{\partial}{\partial x_1} + y_2 \frac{\partial}{\partial y_1} + x_1 \frac{\partial}{\partial x_2} - y_1 \frac{\partial}{\partial y_2}, \\
\tilde{V} &= -y_2 \frac{\partial}{\partial x_1} - x_2 \frac{\partial}{\partial y_1} + y_1 \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_2},
\end{aligned}$$

These vector fields are not tangent to Σ, but we can project out the Liouville direction with the formula

$$U = \tilde{U} - \frac{dH(\tilde{U})}{dH(X)} X, \quad V = \tilde{V} - \frac{dH(\tilde{V})}{dH(X)} X.$$

Lemma 17.1.5. *The form $\lambda = \iota_X \omega$ restricts to the contact form α on Σ. Furthermore, the Reeb vector field of α equals $R := 2X_H$. The vector fields U and V span the contact structure $\xi = \ker \alpha$.*

Proof. The vector field X is Liouville and clearly transverse to Σ since $dH(X) = H$. Hence $\alpha = \iota_X \omega|_\Sigma$ is a contact form. From the remark after Equation (2.14), we see that $X_H \in \ker d\alpha$. We also have

$$\iota_{X_H} \alpha = \frac{1}{2} \sum\nolimits_j a_j (x_j^2 + y_j^2) = \frac{1}{2}.$$

Finally, we have $\alpha(U) = \alpha(V) = 0$, and

$$d\alpha(U, V) = \sum\nolimits_j (x_j^2 + y_j^2) > 0.$$

For convenient computation, note here that $\lambda_x(v) = -\frac{1}{2}\langle ix, v\rangle$, and use that the unit quaternions $1, j, i, k$ are orthogonal. □

The Reeb flow is extremely simple; in complex coordinates $z_j = x_j + iy_j$, this flow is given by

$$\phi_R^t(z_1, z_2) = (e^{ia_1 t} z_1, e^{ia_2 t} z_2).$$

If a_1 and a_2 are rationally independent, positive numbers, then we see that the following orbits are, up to a time shift, the only periodic Reeb orbits,

$$\gamma_1(t) = (e^{ia_1 t}, 0), \quad \text{and} \quad \gamma_2(t) = (0, e^{ia_2 t}).$$

These orbits have periods $T_1 = \frac{2\pi}{a_1}$ and $T_2 = \frac{2\pi}{a_2}$, respectively. Note that γ_1 has a shorter period, or equivalently smaller action, than γ_2 if $a_2 < a_1$. We compute the linearized return map of γ_1 starting at $(1, 0)$ with respect to the basis

$$U_{(1,0)} = \frac{\partial}{\partial x_2}, \quad V_{(1,0)} = \frac{\partial}{\partial y_2}$$

and find

$$d_{(1,0)} \phi_R^{T_1} = \begin{pmatrix} \cos(2\pi \frac{a_2}{a_1}) & -\sin(2\pi \frac{a_2}{a_1}) \\ \sin(2\pi \frac{a_2}{a_1}) & \cos(2\pi \frac{a_2}{a_1}) \end{pmatrix},$$

which clearly has no eigenvalues equal to 1.

We now define a complex structure J_ξ on ξ compatible with $d\alpha$ by putting $J_\xi U = V$ and $J_\xi V = -U$. This is indeed a compatible complex structure since

$$d\alpha(aU + bV, J(aU + bV)) = d\alpha(aU + bV, aV - bU) = (a^2 + b^2) d\alpha(U, V),$$

which is positive if a and b are not both 0. We extend this complex structure to an SFT-like almost complex structure J on $\Sigma \times \mathbb{R}$ by requiring $J\frac{\partial}{\partial t} = R$.

Converting a PDE into an ODE by symmetry

We first choose multipolar coordinates to make the symmetries more apparent, so we put

$$x_1 = r_1\cos(\theta_1), \quad y_1 = r_1\sin(\theta_1), \quad x_2 = r_2\cos(\theta_2), \quad y_2 = r_2\sin(\theta_2).$$

In these polar coordinates, the vector fields R, U, V are given by

$$\begin{aligned}
R &= \sum_j a_j \frac{\partial}{\partial\theta_j},\\
U &= -\frac{a_2r_2(r_1^2+r_2^2)\cos(\theta_1+\theta_2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_1} + \sin(\theta_1+\theta_2)\frac{r_2}{r_1}\frac{\partial}{\partial\theta_1}\\
&\quad + \frac{a_1r_1(r_1^2+r_2^2)\cos(\theta_1+\theta_2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_2} - \sin(\theta_1+\theta_2)\frac{r_1}{r_2}\frac{\partial}{\partial\theta_2},\\
V &= -\frac{a_2r_2(r_1^2+r_2^2)\sin(\theta_1+\theta_2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_1} - \cos(\theta_1+\theta_2)\frac{r_2}{r_1}\frac{\partial}{\partial\theta_1}\\
&\quad + \frac{a_1r_1(r_1^2+r_2^2)\sin(\theta_1+\theta_2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_2} + \cos(\theta_1+\theta_2)\frac{r_1}{r_2}\frac{\partial}{\partial\theta_2}.
\end{aligned}$$

Define the new vector fields

$$\begin{aligned}
P &= -\cos(\theta_1+\theta_2)U - \sin(\theta_1+\theta_2)V\\
&= \frac{a_2r_2(r_1^2+r_2^2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_1} - \frac{a_1r_1(r_1^2+r_2^2)}{\sum_j a_jr_j^2}\frac{\partial}{\partial r_2}\\
&= a_2r_2(r_1^2+r_2^2)\frac{\partial}{\partial r_1} - a_1r_1(r_1^2+r_2^2)\frac{\partial}{\partial r_2},\\
\Theta &= \sin(\theta_1+\theta_2)U - \cos(\theta_1+\theta_2)V\\
&= \frac{r_2}{r_1}\frac{\partial}{\partial\theta_1} - \frac{r_1}{r_2}\frac{\partial}{\partial\theta_2}.
\end{aligned}$$

Here we have used the fact that $\sum_j a_jr_j^2 = 1$ on Σ. With our above choice of J, we see that these vector fields satisfy $JP = \Theta$ and $J\Theta = -P$. Given the circle symmetry, we will make an ansatz for the solution to the Cauchy–Riemann equation $(d\tilde{u})^{0,1} = 0$ using polar coordinates, $(\rho,\vartheta) \mapsto e^{\rho+i\vartheta}$. These coordinates have a singularity as $\rho \to -\infty$ and we will have to take care of this later. For now fix a point (x_0, y_0) on the unit circle. Then our ansatz is

$$\tilde{u}(\rho,\vartheta) = (R_1(\rho)\cos(\vartheta), R_1(\rho)\sin(\vartheta), R_2(\rho)x_0, R_2(\rho)y_0; a(\rho)).$$

The Cauchy–Riemann equation for $\tilde{u}$ in cylindrical coordinates is

$$\frac{\partial\tilde{u}}{\partial\rho} + J\frac{\partial\tilde{u}}{\partial\vartheta} = 0.$$

We note that our ansatz gives $\frac{\partial \tilde{u}}{\partial \vartheta} = \frac{\partial}{\partial \theta_1}$. The latter vector field can be expressed as a linear combination of the Reeb field R and Θ. On Σ we have $\sum_j a_j(x_j^2 + y_j^2) = 1$, and with this we compute

$$\frac{\partial}{\partial \theta_1} = r_1 r_2 \left(\frac{r_1}{r_2} R + a_2 \Theta \right) = r_1^2 R + a_2 r_1 r_2 \Theta.$$

Inserting our ansatz and writing out with our explicit form of J gives

$$\begin{aligned}
0 &= \frac{\partial a(\rho)}{\partial \rho} \frac{\partial}{\partial t} + \frac{\partial R_1(\rho)}{\partial \rho} \frac{\partial}{\partial r_1} + \frac{\partial R_2(\rho)}{\partial \rho} \frac{\partial}{\partial r_2} + J(u(\rho, \vartheta)) \cdot \frac{\partial}{\partial \theta_1} \\
&= \frac{\partial a(\rho)}{\partial \rho} \frac{\partial}{\partial t} + \frac{\partial R_1(\rho)}{\partial \rho} \frac{\partial}{\partial r_1} + \frac{\partial R_2(\rho)}{\partial \rho} \frac{\partial}{\partial r_2} - R_1(\rho)^2 \frac{\partial}{\partial t} \\
&\quad - a_2^2 R_1(\rho) R_2(\rho)^2 (R_1(\rho)^2 + R_2(\rho)^2) \frac{\partial}{\partial r_1} \\
&\quad + a_1 a_2 R_1(\rho)^2 R_2(\rho) (R_1(\rho)^2 + R_2(\rho)^2) \frac{\partial}{\partial r_2}.
\end{aligned}$$

We will bring this into standard ODE form, but first observe that the equation for R_2 can be eliminated due to the relation $H|_\Sigma = \frac{1}{2}$. More precisely, we have

$$R_2(\rho)^2 = \frac{1 - a_1 R_1(\rho)^2}{a_2},$$

which we will plug in into the above. We now also insert $a_1 = 1$, and choose an irrational number $a_2 \in (0, 1)$. That leaves us with the following smooth ODE

$$\begin{aligned}
\frac{d}{d\rho} a(\rho) &= R_1(\rho)^2 \\
\frac{d}{d\rho} R_1(\rho) &= R_1(\rho)(1 - R_1(\rho)^2)((a_2 - 1) R_1(\rho)^2 + 1).
\end{aligned}$$

The equation for R_1 has obvious solutions $R_1 \equiv 0$ and $R_1 \equiv 1$. It follows that for any initial condition $r \in (0, 1)$, the initial value problem $R_1(0) = r$ has a unique solution for all ρ. For later purposes, we will fix $R_1(0) = \frac{1}{2}$. To get an idea of the behavior of the solution, note that the equation for $\frac{d}{d\rho} R_1(\rho)$ shows that the solution $R_1(\rho)$ is strictly increasing. A quick look at the linearization at $R_1 = 0$ and $R_1 = 1$ shows that we have exponential convergence near those points. The function $a(\rho)$ can then be solved for by integration. We have one integration constant, which we can fix by requiring that $a(0) = 0$.

Since polar coordinates are singular near the origin, we now investigate smoothness of the solution as $\rho \to -\infty$ to see whether this solution extends to a map $\mathbb{C} \to \Sigma \times \mathbb{R}$. For $\rho \leq 0$, we have $\frac{d}{d\rho} R_1 \geq \frac{1}{2} R_1$, assuming that a_2 is close to 1. Integrating backwards, we find

$$0 \leq R_1(\rho) \leq \frac{1}{2} e^{\rho/2} \text{ for } \rho \leq 0.$$

This gives us also bounds for $a(\rho)$ by integration,

$$-\frac{1}{4} \leq a(\rho) \leq 0.$$

We now use these bounds to estimate the area of the holomorphic curve $\tilde{u}$ on the punctured unit disk $\dot{D} := \{e^{\rho+i\vartheta} \mid \rho \leq 0\}$. We find for this area

$$\begin{aligned} A(\tilde{u}) &= \int_{\dot{D}} \tilde{u}^*(e^t dt \wedge \alpha + e^t d\alpha) \leq \int_{\dot{D}} \tilde{u}^*(dt \wedge \alpha + d\alpha) \\ &\leq \int_{-\infty}^{0} \int_{0}^{2\pi} (R_1(\rho)^4 + a_2^2 R_1(\rho)^2 R_2(\rho)^2 d\alpha(P(u(\rho,\vartheta)), \Theta(u(\rho,\vartheta)))) d\vartheta d\rho \\ &\leq C \int_{-\infty}^{0} e^{\rho} d\rho = C, \end{aligned}$$

where $C = 2\pi(1 + a_2^2 \max d\alpha(P, \Theta))$. Since this area is finite, Theorem 16.1.1 on the removal of singularities shows that $\tilde{u}$ extends smoothly to a J-holomorphic plane $\tilde{u} : \mathbb{C} \to \Sigma \times \mathbb{R}$. This plane is asymptotic to the non-degenerate periodic Reeb orbit $\gamma_1(t) = (e^{it}, 0)$. We now verify that this plane has finite Hofer energy following formula (13.3). Namely, by Stokes' theorem we find using the properties of our explicit solution

$$\begin{aligned} E(\tilde{u}) &= \sup_{\phi \in \Gamma} \int_{\mathbb{C}} \tilde{u}^* d(\phi\alpha) \\ &= \sup_{\phi \in \Gamma} \left(\lim_{\rho \to \infty} \phi(a(\rho)) \int_{S^1} u(\rho, \cdot)^* \alpha \right) \\ &= 2\pi \sup_{\phi \in \Gamma} \left(\lim_{\rho \to \infty} \phi(a(\rho)) R_1^2(\rho) \right) = 2\pi. \end{aligned}$$

One can directly check that the asymptotic orbit γ_1 has Conley–Zehnder index 3 using the U, V-trivialization, but with all the theory that we have now established, we commit a little mathematical "crime" by using a theorem to prove a lemma. Indeed, we just observe that Σ bounds a convex set, so by Theorem 12.2.1 the contact manifold (Σ, α) is dynamically convex. It follows that γ has a Conley–Zehnder index at least 3. Since the plane and its projection to Σ are embedded, the plane $\tilde{u}$ is a fast plane, and we see that we can indeed apply Theorem 17.1.3 to this example to conclude the proof of Corollary 17.1.4.

Although, we won't need this, let us point out that this plane is also Fredholm regular by Lemma 15.6.1.

17.1.2 Invariant surfaces of section and linking

First let us recall that the hypersurface $\Sigma_{\mu,c}$ is invariant under the anti-symplectic involution $(z_1, z_2) \mapsto (\overline{z}_1, -\overline{z}_2)$ on $\mathbb{C}^2$ and the retrograde periodic orbit is symmetric with respect to this anti-symplectic involution. As a stronger corollary than the previous one we deduce now the existence of an invariant global surface of section, which is Theorem 9.5.3.

Corollary 17.1.6. *Under the assumptions of Corollary* 17.1.4 *there exists an invariant global surface of section bounded by the retrograde periodic orbit.*

Proof. Because the space of SFT-like complex structures is connected, we can homotope any SFT-like complex structure to an anti-invariant SFT-like complex structure, i.e., an SFT-like complex structure J which satisfies

$$\rho^* J = -J$$

for the anti-symplectic involution ρ. By Theorem 17.1.3 we can now find a fast finite energy plane of vanishing Siefring self-intersection number asymptotic to the retrograde periodic orbit with respect to the anti-invariant SFT-like complex structure J. Because J is anti-invariant, the anti-symplectic involution ρ induces an involution

$$\rho_* \colon \widehat{\mathcal{M}}_{\text{fast}}(\gamma_R) \to \widehat{\mathcal{M}}_{\text{fast}}(\gamma_R)$$

given for $\widetilde{u} \in \widehat{\mathcal{M}}_{\text{fast}}(\gamma_R)$ by

$$\rho_* \widetilde{u}(z) = \rho(\widetilde{u}(\overline{z})), \quad z \in \mathbb{C}.$$

Consider the open book decomposition as described in the proof of Theorem 17.1.1. Pick a point

$$x \in \mathrm{Fix}(\rho) \cap \big(\Sigma_{\mu,c} \setminus \mathrm{im}(\gamma_R)\big)$$

and let $\widetilde{u} = (u, a)$ be the fast finite energy plane asymptotic to γ_R such that

$$x \in \mathrm{im}(u).$$

Because x lies in the fixed point set of ρ it follows that

$$x \in \mathrm{im}(\rho_* u).$$

Therefore the pages $\mathrm{im}(u)$ and $\mathrm{im}(\rho_* u)$ of the open book decomposition contain the common point x. However, two pages are either disjoint or coincide. Therefore

$$\mathrm{im}(u) = \mathrm{im}(\rho_* u)$$

and $\mathrm{im}(u)$ is an invariant global surface of section. □

Theorem 17.1.7. *Assume that $\Sigma \subset \mathbb{C}^2$ is a dynamically convex starshaped hypersurface and $\gamma \in C^\infty(S^1, \Sigma)$ is a simple non-degenerate periodic Reeb orbit for which there exists a fast finite energy plane $\widetilde{u} = (u, a)$ with vanishing Siefring self-intersection number asymptotic to γ. Then γ is linked to every other Reeb orbit on Σ different from itself.*

Proof. First assume that Σ is non-degenerate in the sense that all periodic Reeb orbits on Σ are non-degenerate. Then it follows from Theorem 16.4.5 that the moduli space $\mathcal{M}_{\text{fast}}(\gamma)/\mathbb{R}$ is compact. In particular, the same proof as in Theorem 17.1.1 shows that $\text{im}(u)$ is a global surface of section. In particular, every periodic orbit different from γ is linked to γ.

Now we drop the assumption that Σ is non-degenerate. Assume that γ' is a periodic Reeb orbit on Σ different from γ. Now perturb Σ slightly to make it non-degenerate in such a way that γ and γ' still survive. By definition of the Conley–Zehnder index in the degenerate case the Conley–Zehnder index can only increase under a small perturbation. Therefore we can assume that after perturbation all periodic orbits of period less than the period of γ still have Conley–Zehnder index greater or equal to three. Now the argument from the non-degenerate case shows that γ and γ' are linked. □

Remark 17.1.8. Theorem 17.1.7 immediately implies Theorem 8.3.9 in Chapter 8.

17.1.3 Global surface of section to open book

Proposition 17.1.9. *Suppose that we are given a contact form α on S^3 together with a global disk-like surface of section $D : D^2 \to S^3$ for the Reeb flow. Denote by $\gamma := D|_{\partial D^2} : S^1 \to S^3$ the periodic orbit that bounds D and assume that*

- *the orbit γ is transversely non-degenerate, and*
- *the Conley–Zehnder index satisfies $\mu_{CZ}(\gamma) \geq 3$.*

Then there is an open book on S^3 with binding γ whose pages are transverse to R. Furthermore, the return map extends continuously to the boundary as a homeomorphism of D^2 that is conjugate to a map preserving the $D^* d\alpha$-area.

Proof. First of all, we need exhibit for every point $p \in \text{int}(D^2)$ a return time $t(p)$ depending smoothly on p such that the return map $p \mapsto D^{-1} \circ \phi_R^{t(p)} \circ D(p)$ is smooth on the *interior* of the disk D^2. It is clear that this holds by smooth dependence on initial conditions. Define the return map

$$\begin{aligned} rt : D^2 &\longrightarrow D^2 \\ x &\longmapsto D^{-1} \circ \phi_R^{t(x)} \circ D(x), \end{aligned}$$

where we will define $t(x)$ for points in the boundary below.

Step 1. Finding a good neighborhood: We claim that there is a neighborhood $\nu(\gamma)$ of γ together with coordinates $\psi : S^1 \times D^2 \to \nu(\gamma)$ such that

$$\psi^{-1}(\nu(\gamma) \cap \text{im}\, D\,) = \{(\vartheta; x, 0) \in S^1 \times D^2 \mid \vartheta \in S^1, \text{ and } x \geq 0\},$$

and such that $d_{(\vartheta,0,0)}\psi \frac{\partial}{\partial x}$ is tangent to the contact structure ξ.

To see this, we adapt the tubular neighborhood theorem a little. Choose a trivialization of the contact structure along γ, so a map $S^1 \times \mathbb{R}^2 \to \gamma^*\xi$ with the property that the vector X, which we define as the image of the vector $(1, 0)$, is tangent to D along γ and pointing inward. Let Y denote the image of the vector $(0,1)$. Note that such a trivialization does *not* extend to a trivialization of the contact structure along the disk D; the winding number of the loop of symplectomorphisms needed to transform (X,Y) into a frame that extends over the disk equals -1 (namely going from polar to Cartesian coordinates).

Choose a metric on S^3 such that $\operatorname{im} D$ is a geodesic submanifold. For example, we may choose any metric g_D on $\operatorname{im} D$, and extend this by $g_D + dn \otimes dn$, where n is a coordinate for the complement of the disk. We define the desired map ψ as

$$\psi : S^1 \times D^2 \longrightarrow \nu(\gamma), \quad (\vartheta; x, y) \longmapsto \exp_{D(e^{i\vartheta})}(xX + yY).$$

The projection $\pi : S^1 \times (D^2 \setminus \{0\}) \to S^1, (\vartheta; z) \mapsto \frac{z}{|z|}$ will serve as the analog of the projection in an open book.

Put $y = 0$ and consider the curve $s \mapsto \exp_{D(e^{i\vartheta})}(sX)$, and apply the map ϕ_R^t to this curve. The derivative at $s = 0$ is the time-t linearized flow of the Reeb field R acting on X, or in a formula

$$d\phi_R^t(X_{D(e^{i\vartheta})}).$$

Step 2. Extending the return map to the boundary: Let θ denote the rotation number of γ with respect to a trivialization of $D^*\xi$. By Theorem 10.6.1, the Conley–Zehnder index $\mu_{CZ}(\gamma)$ equals $2\lfloor\theta\rfloor + 1$ if γ is elliptic and 2θ if γ is hyperbolic. As $\mu_{CZ}(\gamma) \geq 3$, we see that $\theta > 1$. This implies that the rotation number with respect to the (X,Y)-trivialization is greater than 0.

We look at the flow using the tubular neighborhood we constructed, so we consider the map

$$\psi^{-1} \circ \phi_R^t \circ \exp_{D(e^{i\vartheta})}(sX).$$

Taking the derivative at $s = 0$ gives as a bundle map from $S^1 \times \mathbb{R}^2 \to S^1 \times \mathbb{R}^2$, sending $(t_0; z) \mapsto (t_0 + t; [d\phi_R^t] \cdot (1,0))$, where $[d\phi_R^t]$ is the path of symplectic matrices with respect to the (X,Y)-trivialization. Since the rotation number with respect to the (X,Y)-trivialization is positive, the path $[d\phi_R^t]$ rotates in positive sense, and so we can find a unique minimal, positive $t_{ret} = t_{ret}(\vartheta)$ such that $[d\phi_R^{t_{ret}}] \cdot (1,0) \in \mathbb{R}_{>0} \times \mathbb{R}$.

This gives us the definition of $t(x) = t_{ret}(\vartheta)$ for points x on the boundary, so we have now defined the map rt on the whole disk D^2. The return map at the boundary is the smooth map $\vartheta \mapsto \vartheta + t_{ret}(\vartheta)$. We conclude that rt is a homeomorphism of D^2 that is smooth on the interior with the property that it is conjugate to a map preserving the $D^*d\alpha$-area.

The open book part now follows quickly. Note that $S^3 \setminus \gamma(S^1) \cong \mathring{D}^2 \times \mathbb{R}/(z,s) \sim (rt(z), s - t(z))$ is clearly a bundle over the circle, and near the binding the map π can be adapted giving us an open book whose pages are transverse to the Reeb vector field. $\square$

We point out that the return map, which is not equal to the identity near γ, is not equal to the so-called monodromy; we will describe this notion later in Section 17.1.4.

Remark 17.1.10. The non-degeneracy is not essential. Given a degenerate binding, one can obtain the same result as long as the Conley–Zehnder index is at least 3 (or more directly as long as the rotation number is larger than 1). On the other hand, this condition on the Conley–Zehnder index is necessary. The example below, which we have learned from Pedro Salomão, shows this.

For easy computations, we first describe the situation for the standard Hopf contact form $\alpha = \iota^* \frac{i}{2} \sum_j (z_j d\bar{z}_j - \bar{z}_j dz_j)$ on $\iota : S^3 \to \mathbb{C}^2$. Consider the map

$$T : S^1 \times D^2 \longrightarrow S^3, \quad (z; r, \theta) \longmapsto (\sqrt{1-r^2}e^{iz}, re^{i\theta}).$$

This pulls back the Hopf form to

$$T^*\alpha = (1-r^2)dz + r^2 d\theta,$$

and the Reeb field is $R_T = \partial_z + \partial_\theta$. Since the flow is periodic, each orbit is the binding orbit of a global surface of section. We will take the disk

$$D : D^2 \longrightarrow S^3, \quad (\rho, \phi) \longmapsto (\rho e^{i\phi}, \sqrt{1-\rho^2}),$$

which is transverse to the flow. The return map is the identity, which obviously extends continuously to the boundary.

We now modify the contact form on $S^1 \times D^2$ into

$$\tilde{\alpha} = h_1(r)dz + h_2(r)d\theta,$$

where h_1 and h_2 are profile functions satisfying the following conditions.

1. Near $r = 1$, we have $h_1(r) = 1 - r^2$ and $h_2(r) = r^2$.
2. h_1 is decreasing, and near $r = 0$, we have $h_2(r) = r^2$.
3. $h(r) := h_1(r)h_2'(r) - h_2(r)h_1'(r) > 0$ for $r > 0$.

By writing out $\tilde{\alpha} \wedge d\tilde{\alpha} > 0$, we see that the third condition is equivalent to the contact condition. Condition (1) is imposed to make $\tilde{\alpha}$ coincide with the Hopf form near $r = 1$ to ensure a nice extension to S^3. The Reeb vector field of $\tilde{\alpha}$ is given by

$$\tilde{R} = \frac{h_2'(r)\partial_z - h_1'(r)\partial_\theta}{h(r)}.$$

We hence see that D still gives a global disk-like surface of section provided that $h_1'(r) < 0$. The return time is $2\pi \frac{h(r)}{-h_1'(r)}$, so the return map is given by

$$\psi : D^2 \longrightarrow D^2, \quad (\rho, \phi) \longmapsto \left(\rho, \phi - 2\pi \frac{h_2'(r)}{h_1'(r)}\right).$$

If we choose $h_1(r) = 1 - r^4$ near $r = 0$, then we see that the map ψ does not extend continuously to the boundary, since angle ϕ is rotated infinitely fast at the boundary. It follows that the Conley–Zehnder index of the binding orbit at $r = 0$ is less than 3.

Remark 17.1.11. The choice $h_1(r) = 1 - r^4$ and $h_2(r) = r^2$ can be realized by restricting $\frac{i}{2}\sum_j (z_j d\bar{z}_j - \bar{z}_j dz_j)$ to the set

$$|z_1|^2 + |z_2|^4 = 1.$$

Namely, we simply parametrize by $z_1 = \sqrt{1-r^4}e^{i\theta}$ and $z_2 = re^{i\phi}$ to obtain the form $\tilde{\alpha}$. Note that this hypersurface bounds, in fact, a convex, but not strictly convex set.

For completeness, we briefly compute the Conley–Zehnder index of the *degenerate* periodic orbit at $r = 0$. To do this, we first note that the following frame is globally defined, and provides a symplectic trivialization ϵ_{UV} of $(\ker\tilde{\alpha}, d\tilde{\alpha})$,

$$U = \sqrt{\frac{h_2(r)}{h_1(r)}}\sin(z+\phi)\partial_z + \sqrt{h_1(r)}\cos(z+\phi)\partial_r - \sqrt{\frac{h_1(r)}{h_2(r)}}\sin(z+\phi)\partial_\phi,$$

and

$$V = -\sqrt{\frac{h_2(r)}{h_1(r)}}\cos(z+\phi)\partial_z + \sqrt{h_1(r)}\sin(z+\phi)\partial_r + \sqrt{\frac{h_1(r)}{h_2(r)}}\cos(z+\phi)\partial_\phi.$$

To obtain this mysterious looking frame, we have started with the standard frame ix, jx, kx on S^3, where $x \in S^3$ and i, j, k are the standard quaternions. The vector field $x \mapsto ix$ is the Reeb vector field, and $x \mapsto jx$, $x \mapsto kx$ trivialize the standard contact structure. We then pulled this frame back using T, and modified it to accommodate for our perturbed contact form.

With this in mind, we claim that ϵ_{UV} is a global trivialization of the contact structure on S^3, and in fact there is only one such trivialization up to homotopy, since $\pi_3(Sp(1)) = 0$. On the other hand, along the orbit $\gamma : \mathbb{R}/2\pi\mathbb{Z} \to S^1 \times D^2$, $t \mapsto (t, 0, 0)$, we also have the frame ϵ_{XY} consisting of

$$X = \frac{\partial}{\partial x} = \cos(\phi)\frac{\partial}{\partial r} - \sin(\phi)\frac{1}{r}\frac{\partial}{\partial r}, \quad Y = \frac{\partial}{\partial y} = \sin(\phi)\frac{\partial}{\partial r} + \cos(\phi)\frac{1}{r}\frac{\partial}{\partial r}.$$

At $r = 0$, a short computation shows that the two trivializations are related by

$$U = \cos(z)X - \sin(z)Y, \quad V = \sin(z)X + \cos(z)Y.$$

Since the Maslov index (or winding number) of the loop

$$t \mapsto \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}$$

equals 1, we find

$$\mu_{CZ}(\gamma,\epsilon_{UV}) = 2\mu(\begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}) + \mu_{CZ}(\gamma,\epsilon_{XY}) = 2 + \mu_{CZ}(\Psi_S),$$

where Ψ_S is the path of the symplectic matrices corresponding to the operator $-J\partial_t - S(t)$. Since the linearized flow of $\tilde{R}$ with respect to the trivialization ϵ_{XY} is constant, we see with formula (11.2) that $S(t) \equiv 0$, so we conclude that $\mu_{CZ}(\gamma,\epsilon_{UV}) = 2$.

17.1.4 Topological restrictions on open books

It is known by an old result of Alexander [12] that every closed, orientable 3-manifold admits an open book. This result is purely topological and gives no dynamical information. On the other hand, given the page of an open book there are restrictions on the possible topologies that can be obtained. This directly follows from the definition, and the following construction. Choose a Riemannian metric on Σ for which each fiber of θ is orthogonal to the vector field $d\epsilon\frac{\partial}{\partial\phi}$, which we defined using the trivialization ϵ. Extend the metric in any way, and lift the vector field $\frac{\partial}{\partial\phi}$ to a vector field X_θ that is orthogonal to the fibers of θ. By our choice of metric near B, the vector field X_θ coincides with $d\epsilon\frac{\partial}{\partial\phi}$. Define $\psi_t := \phi^t_{X_\theta}$. Then we have

$$\Sigma \setminus B \cong P \times \mathbb{R}/(p,t) \sim (\psi_1(p), t-1).$$

Remark 17.1.12. The map ψ_1 is called the *monodromy* of the open book, and is well defined up to isotopy and conjugacy. This notion is not to be confused with the return map which is similar, but different since the return map has no control on its behavior near the boundary, whereas the monodromy is always the identity near the boundary.

Since the monodromy is by construction the identity near the boundary, we obtain the following decomposition for Σ,

$$\Sigma = B \times D^2 \cup_{\mathrm{id}_{B\times S^1}} P \times [0,1]/(\psi_1(p), t-1),$$

where we view B as the boundary of P.

The monodromy and topological type of the page determine the whole manifold as the following lemma shows.

Lemma 17.1.13. *Suppose that Σ_i has open book (B_i,θ_i) for $i = 0,1$. If the pages of these open books are homeomorphic and their monodromies are isotopic relative to the boundary (up to conjugacy by a homeomorphism of the pages), then Σ_0 is homeomorphic to Σ_1.*

Proof. We directly identify the fibers of θ_i and denote them by P. Hence we find

$$\Sigma_i \cong \partial P \times D^2 \cup_{\mathrm{id}} P \times [0,1]/\sim_i,$$

where $(p, 1) \sim_i (h_i(p), 0)$ for $i = 0, 1$. The equivalence classes will be denoted by $[p, t]_i$. By assumption we find an isotopy $\{h_s\}_{s \in [0,1]}$ between h_0 and h_1 that is the identity on the boundary of P. Define the map

$$f := P \times [0,1]/\sim_0 \longrightarrow P \times [0,1]/\sim_1, \quad [p,t]_0 \longmapsto [h_{1-t} \circ h_0^{-1}(p), t]_1.$$

This map is well defined, since

$$f([p,1]_0) = [p,1]_1 = [h_1 \circ h_0^{-1} \circ h_0(p), 0]_1 = f([h_0(p), 0]_0).$$

This map is also continuous and invertible. Furthermore, f can be extended to $\partial P \times D^2$ as the identity. It follows that f is a homeomorphism. □

Since we are using very specific holomorphic curves, namely finite energy planes, to construct global surfaces of section in this monograph, we briefly point out the following result.

Proposition 17.1.14. *Suppose that Σ is a 3-manifold with an open book whose page is a disk. Then Σ is homeomorphic to S^3.*

Proof. By assumption Σ has an open book decomposition whose page is a disk. The monodromy h is the identity in a neighborhood of the boundary, and is hence orientation preserving, so $h \in Homeo_c^+(D^2)$. The so-called Alexander trick now shows that h is isotopic to the identity. Indeed,

$$h_t(p) = \begin{cases} t \cdot h(\frac{p}{t}) & \|x\| < t \\ \mathrm{id} & t \leq \|p\| \leq 1 \end{cases}$$

provides a continuous isotopy between h at $t = 1$ and the identity at $t = 0$. By Lemma 17.1.13 and the example in Section 9.4.1 we conclude that Σ is homeomorphic to S^3. □

Remark 17.1.15. Another case of interest is when the page is an annulus. In that case Σ can shown to be homeomorphic to a lens space.

Chapter 18

Numerics and Dynamics via Global Surfaces of Section

In this chapter we will see more practical applications of global surfaces of section to the restricted three-body problem. This involves a lot of numerical work, which we will briefly outline. We will also give a rough overview of periodic orbits that appear together with some pictures to get a better idea of the dynamics, but we will omit full proofs, which often require some computer assisted arguments. The nature of this chapter is therefore quite different from the rest of this book.

Needless to say, there is a vast literature on this subject. We will mainly look at periodic orbits that appear below the first critical value. The more interesting orbits will be left out. These orbits tend to appear above the first critical value. See the work of Hénon and Szebehely for more on this topic, [115, 231].

Unless otherwise mentioned, we will write H for the Hamiltonian of the restricted three-body problem after Levi-Civita regularization, and we will only consider orbits after regularization. However, we draw pictures of the orbits in the unregularized coordinates using the Levi-Civita map

$$\big(Q(z,w), P(z,w)\big) = \left(2z^2, \frac{w}{\bar{z}}\right). \tag{18.1}$$

18.1 Symmetric orbits

We start with a discussion of symmetric orbits in the rotating Kepler problem. The figures have been obtained by numerical integration, but we want to point out that the rotating Kepler problem can be and has been analyzed by analytical methods. For instance, we can apply Lemma 8.2.1 to reproduce the figures. The study of the rotating Kepler problem goes back to even well before Poincaré.

The following orbits, Hecuba $(2,1)$, Hilda $(3,2)$ and Thule $(4,3)$ are orbits that come out of a simple cover of the direct circular orbit in $\mathbb{R}P^3$. These orbits all have linking number 1 with the retrograde orbit.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8_18

The orbits are drawn for $\mu = 0$, and hence they come in S^1-families with the corresponding Morse–Bott non-degeneracy, see Section 7.3 for the definition. In particular, these orbits are degenerate. However, we have already seen in Section 8.4 that they are non-degenerate as symmetric orbits. Hence there will be corresponding orbits for small, positive μ. The resulting orbits are then often, but not necessarily, non-degenerate as periodic orbits without the adjective symmetric [5]. If this happens, these perturbed orbits will be elliptic or hyperbolic. We have indicated the adjectives "elliptic" and "hyperbolic" to indicate what kind of non-degenerate orbit they become after perturbation to $\mu > 0$.

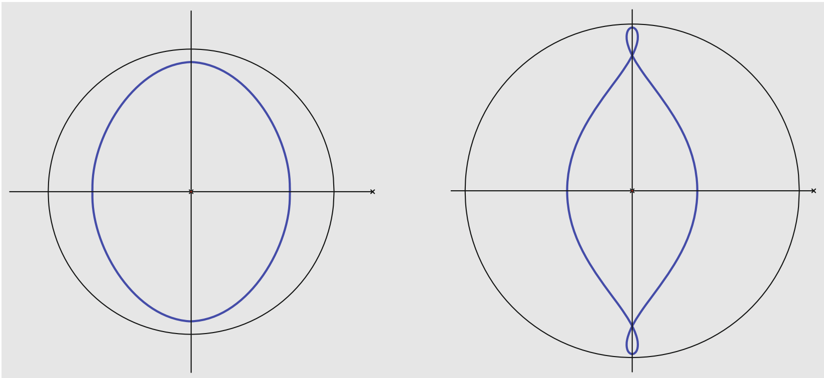


Figure 18.1: An "elliptic" Hecuba just after birth ($\mu = 0, c = 1.58$) and later in life ($\mu = 0, c = 1.51$).

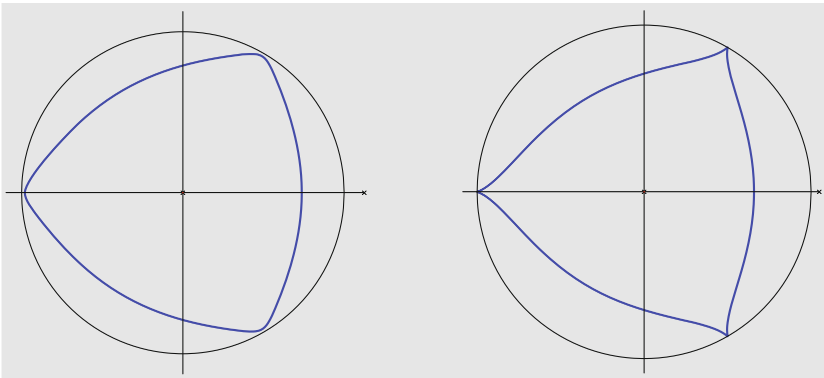


Figure 18.2: An "elliptic" Hilda just after birth ($\mu = 0, c = 1.52$) and later in life ($\mu = 0, c = 1.51$).

[5]This requires either some numerical work or an additional argument based for example on the averaging method, which we briefly discussed in Section 8.2.3.

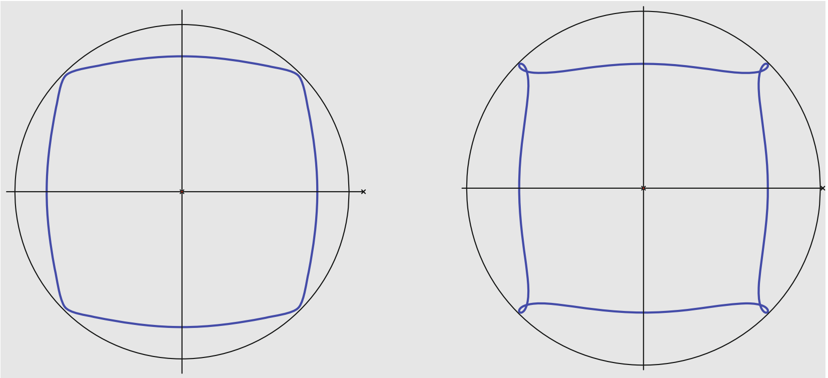


Figure 18.3: An "elliptic" Thule just after birth ($\mu = 0, c = 1.51$) and later in life ($\mu = 0, c = 1.501$).

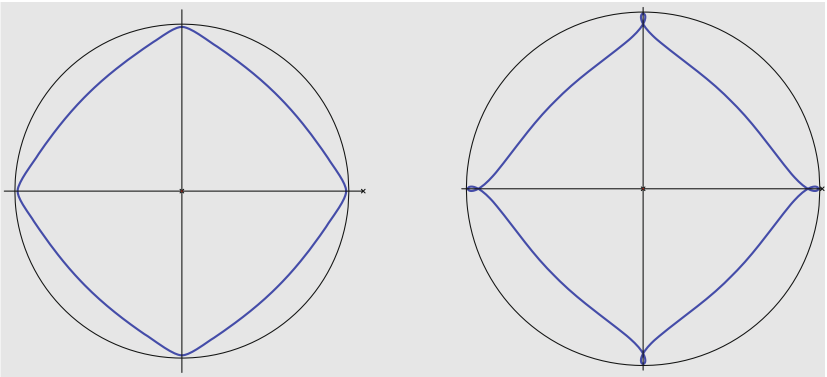


Figure 18.4: A "hyperbolic" Thule just after birth ($\mu = 0, c = 1.51$) and later in life ($\mu = 0, c = 1.501$).

18.2 Finding orbits via shooting

By shooting we can follow periodic orbits by varying the parameters. This is most easily done for symmetric orbits, since we can use one-dimensional surfaces of section. The intermediate value theorem, together with error bounds on the solutions obtained by numerical integration, will guarantee the existence of periodic orbits in a simple way. One way of doing this is to combine the Taylor method for numerical integration with interval arithmetic. We will give a brief outline below.

In the following we consider symmetric orbits with two axis crossings. To make precise what we are talking about, we say that a periodic orbit (γ, τ) has k *axis crossings* if

- the projection to the q-plane, $\pi_q \circ \gamma$, intersects the x-axis transversely
- $(\pi_q \circ \gamma)^{-1}(\mathbb{R} \times \{0\})$ consists of k points.

This notion is convenient for numerical integration, but it is not an invariant of a family of (non-degenerate) periodic orbits. Axis crossings can get lost in a collision or at the boundary of the Hill's region, as we can see in Figure 18.4.

We point out that shooting is a direct adaptation of Birkhoff's proof of the existence of a retrograde orbit, but it can be used to find any other symmetric orbit. Of course, we are not making any claims to originality here. This method was already used by Hénon in 1969 and others before him, see for example [114]. We also point out that there are non-symmetric periodic orbits. These can also be found by shooting type procedures, but one needs more than a one-dimensional surface of section.

The main reason for the emphasis on symmetric periodic orbits is that periodic orbits below the first Lagrange value with small action seem to be often symmetric. Most importantly, the retrograde orbit is. Although no one sofar has managed to give an analytic proof of the existence of the direct orbit, numerical shooting shows that there is often such an orbit. In Figure 18.1 we see that the circular direct orbit becomes an elliptic Hecuba for $\mu > 0$. For larger values of μ and small values of the Jacobi energy c there may not be a "circular" direct orbit.

18.2.1 Following an orbit by varying μ

We can keep c constant and vary μ. Unsurprisingly, such a 1-parameter family of periodic orbits can have bifurcations, just like in the case when one varies c. One point one has to keep in mind is that one can pass the Lagrange point: for $\mu = 0$ the first critical value is $3/2$, but for $\mu = 1/2$ the first critical value equals 2. As a result many familiar orbits from the rotating Kepler problem do not appear for other mass ratios below the first critical value. In Figure 18.5 we see an example with Hestia, which becomes degenerate for $c = 1.66$ and varying μ from 0 to 0.07.

18.2.2 Conley–Zehnder index

By considering the linearized differential equations, we can also numerically approximate the linearized flow. In particular, we obtain an approximation to the path of symplectic matrices needed for the computation of the Conley–Zehnder index of a periodic orbit. For practical purposes, the rotation number formula for the Conley–Zehnder index from Theorem 10.6.1 is convenient. This formula works for *non-degenerate* periodic orbits, which are in practice the only periodic orbits that can be found by shooting.

We used this to compute the Conley–Zehnder indices of periodic orbits. We do not have a complete argument, so we cannot call it a theorem, but it seems that the retrograde periodic orbit is the only orbit with Conley–Zehnder index 3 in the Levi-Civita regularized restricted three-body problem for all energies below

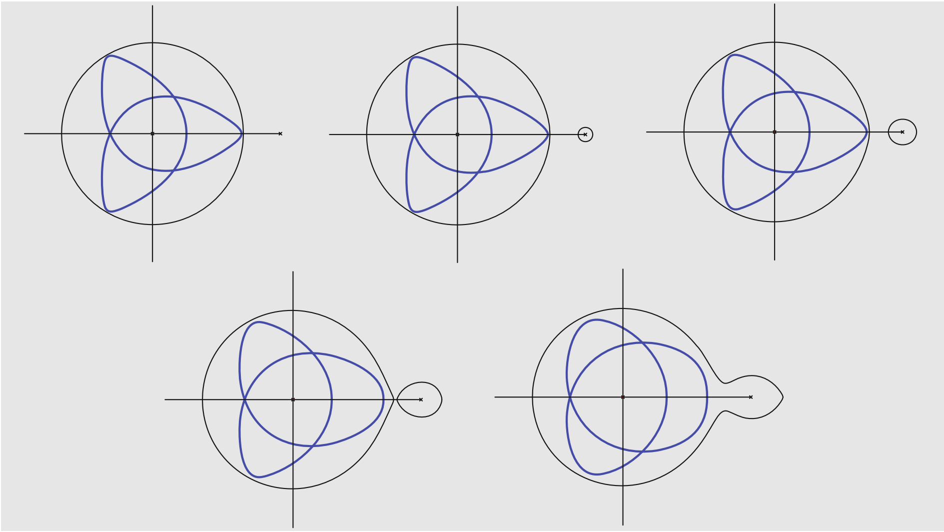


Figure 18.5: A hyperbolic Hestia with varying μ.

the first critical value. Other orbits seem to have both larger action and index. In other words, it appears that the restricted three-body problem is dynamically convex.

We point out that this approach is unlikely to prove dynamical convexity. The problem is that there are (probably) infinitely many, geometrically distinct periodic orbits for each energy and mass-ratio, and this direct approach only checks the index for finitely many families of orbits.

18.2.3 How to make the numerics rigorous

A good reference to topic of validated numerics is the monograph by Tucker, [235]. Here we give a brief outline of some aspects of this topic for an analytic ODE $\dot{x} = f(x)$ in $\mathbb{R}^n$ with initial value $x(0)$. We first make the brief observation that we can, in principle, write down the power series solution to any order by the formula

$$x(t) = x(0) + \sum_{j=1}^{\infty} \frac{1}{j!} \underbrace{\nabla_f \nabla_f \ldots \nabla_f}_{j \text{ terms}} f(x(0)) t^j. \tag{18.2}$$

Indeed, taking the derivative of the equation, we find $\ddot{x} = \nabla_{\dot{x}} f = \nabla_f f$, so by induction we see that the nth derivative satisfies

$$x^{(n)}(0) = \nabla_f \nabla_f \ldots \nabla_f f(x(0)).$$

As the equation is analytic, the solution is analytic, too, by the Cauchy Existence Theorem on page 15–16 from [188]. In other words, this power series converges on some disk $|t| < \epsilon$. The truncated Taylor polynomial can then be used as an approximation with an estimate of the remainder.

Remark 18.2.1. Although it is possible to symbolically obtain a truncation of the power series (18.2) by hand or with a computer algebra program, this is usually a bad idea. The reason is that the number of terms in a symbolic expansion often grows too quickly with the degree of the truncation. The idea of *automatic differentiation* can be used deal with this. This is also explained in [235].

18.2.4 Integration with error bounds

One way to obtain rigorous error bounds on a computer is by using interval arithmetic. Books have been written about this topic, so this one paragraph introduction is not going to do the subject justice. We refer to [183] for a real introduction to this topic. The idea is compute with sets rather than approximations to real numbers. Theoretically, this is done by working with the ranges of functions and operations, e.g., $\sin([0,\pi]) = [0,1]$. When implemented on a computer, care has to be taken to always produce an interval enclosing the result of the computation. For example, $\widetilde{\sin}([\tilde{0},\tilde{\pi}])$ should return an interval enclosing $[0,1]$, and because π cannot be represented exactly, the interval $[0,\pi]$ should first be rounded outward to $[\tilde{0},\tilde{\pi}]$, an interval containing $[0,\pi]$, and then be evaluated with a suitable approximation to the sine function. This approximation should return an interval containing the true range; for a general function, this is *not* done by just taking the left and right value, and this has to be done with care.

The second idea is to use the truncated Taylor polynomial of some degree N for (18.2) together with a formula for the remainder such as the Lagrange form. In the Lagrange form for the Taylor polynomial on $[0,h]$, we need an unknown parameter $\tau \in (0,h)$. We will see that interval arithmetic can be used to obtain bounds on τ and by extension, on the remainder term.

Consider an interval vector $[v]$ in $\mathbb{R}^n$. This is a set of the form $[v] = [a_1,b_1] \times \cdots \times [a_n,b_n]$. We may think of this interval vector as a set enclosing the point $x(0) = (x_1,\ldots,x_n) \in \mathbb{R}^n$ which we are interested in; we have $x_i \in [a_i,b_i]$, and call $[v]$ an *enclosure* of $x(0)$. We first find an a priori enclosure of the solution $\phi_f^t(x(0))$, known as a rough enclosure. In general, this can only be found if t is sufficiently small.

Lemma 18.2.2 (Moore–Lohner rough enclosure)**.** *Suppose* $[u_0]$ *and* $[u]$ *are interval vectors such that*

$$[u_1] := [u] + [0,h]f([u_0]) \subset [u_0]$$

Then for every initial value $x_0 \in [u]$*, the solution* $x(t)$ *to the initial value problem*

$$\begin{aligned}\dot{x} &= f(x)\\ x(0) &= x_0\end{aligned}$$

exists on $[0,h]$ *and is contained in* $[u_1]$.

By the lemma, see [167], we know that the true solution is contained in $[u_1] \subset [u_0]$, so we can obtain bounds on the remainder using the rough enclosure $[u_1]$ or even $[u_0]$. As the step-sizes are chosen to be small, and the order can be made arbitrarily high, the error can be made arbitrarily small if we assume exact arithmetic. With interval arithmetic, we therefore obtain good control on the error. This gives us a tight enclosure of the desired solution $\phi_f^h(x(0))$.

This naive method has been known at least since Moore in 1965. It has unfortunately serious drawbacks, since the enclosure tends to grow very quickly when applied several times in succession. However, since the late 1980s Lohner, [167], and many after him have made improvements to make this idea practical and extremely useful.

18.2.5 Finding periodic orbits

The final step is to use this enclosed integration scheme to verify whether enclosures of initial conditions contain periodic orbits. Shooting for symmetric periodic orbits provides the easiest situation. We can simply obtain enclosures of the return map for initial conditions slightly to the left and slightly to the right of the suspected periodic orbit. If we find a sign change (now with enclosure), we are sure that there is a periodic orbit.

In practice, more sophisticated ideas are better. We compute enclosures of the return map including derivatives, and appeal to a degree argument to conclude that there is (or is no) periodic orbit.

18.3 Numerical construction of a foliation by global surfaces of section

To actually approximate global surfaces of section numerically for practical use, we employ two methods:

1. We will describe a numerical neck-stretching construction which is closely related to holomorphic curves.
2. Inspired by Birkhoff we give a conjectural formula for a foliation that is transverse to the flow. This is completely ad hoc.

18.3.1 Numerical holomorphic curves

The stretching construction from SFT can be adapted numerically. We will describe this in a very simple case, but the numerical adaptation requires us to be completely explicit in the construction. The idea is to embed the energy hypersurface $\Sigma_{\mu,c}$ of interest into $\mathbb{C}P^2$: Endow the symplectization neighborhood $(\,(-R,R)\times\Sigma_{\mu,c}, de^t\lambda_{\mu,c}\,)$ with an SFT-like almost complex structure $J_{\mu,c,R}$, and embed this symplectic manifold into a sufficiently large Darboux ball $(D^4_{f(R)}, \omega_0)$. This Darboux ball embeds in turn into $(\mathbb{C}P^2, f(R)\omega_{FS})$, i.e., the complex projective plane with a sufficiently large multiple of the standard Fubini–Study symplectic form. Now extend the almost complex structure $J_{\mu,c,R}$, which we continue to denote by $J_{\mu,c,R}$, to $(\mathbb{C}P^2, f(R)\omega_{FS})$ such that

- $J_{\mu,c,R}$ is standard on a small Darboux ball $D^4_\epsilon \subset D^4_{f(R)}$.
- $J_{\mu,c,R}$ is standard on a neighborhood of the line at infinity $\mathbb{C}P^1_\infty := \{[x : y : 0] \mid [x : y] \in \mathbb{C}P^1\}$.

For a numerical implementation, we need an explicit SFT-like complex structure. Denote the standard Liouville form on D^4 by λ and we can trivialize the contact structure on $\Sigma_{\mu,c}$ by using

$$\xi = \operatorname{span}\left(\tilde{U} = j\nabla H_s - \frac{\lambda(j\nabla H_s)}{\lambda(i\nabla H_s)} i\nabla H_s, \tilde{V} = k\nabla H_s - \frac{\lambda(k\nabla H_s)}{\lambda(i\nabla H_s)} i\nabla H_s\right),$$

where i, j, k are the standard imaginary quaternions satisfying $i^2 = j^2 = k^2 = ijk = -1$. We directly see that $\tilde{U}$ and $\tilde{V}$ lie both in the kernel of dH, so they are tangent to $H^{-1}_{\mu,c}(0)$, and in the kernel of λ, i.e., $\tilde{U}, \tilde{V} \in \ker \lambda|_{H^{-1}_{\mu,c}(0)}$. We have $d\lambda(\tilde{U}, \tilde{V}) \neq 0$, so after rescaling we obtain $U, V \in \ker \lambda|_{H^{-1}_{\mu,c}(0)}$ forming a symplectic basis. Choose a compatible complex structure $J_{\xi,\mu,c}$ for the contact structure ξ by putting $J_{\xi,\mu,c}U = V$, and $J_{\xi,\mu,c}V = -U$.

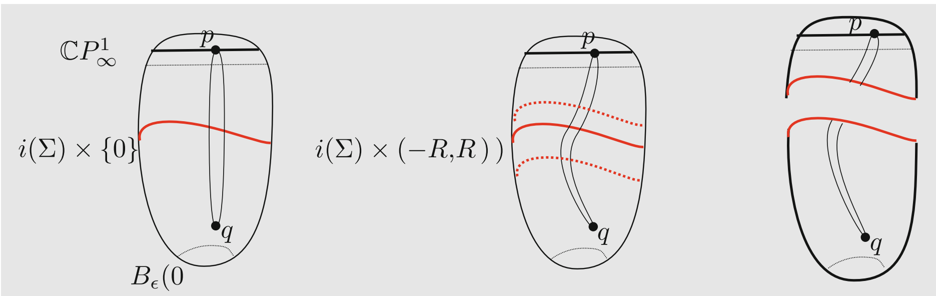


Figure 18.6: Breaking of curves.

Take a point $p \in \mathbb{C}P^1_\infty$ and a point q in the symplectization neighborhood, slightly outside $B_\epsilon(0)$. For $R = 0$, the almost complex structure $J_{\mu,c,0}$ is the standard complex structure, so we can explicitly write down the holomorphic curve as the standard projective line

$$\begin{aligned} u_0 : \mathbb{C}P^1 &\longrightarrow \mathbb{C}P^2 \\ [\lambda : \mu] &\longmapsto [\lambda p + \mu q]. \end{aligned}$$

We obtain a homotopy of compatible almost complex structures $\{J_{\mu,c,R}\}$ by varying the R-parameter. Discretize this homotopy by taking a sequence of R_i's. We find a sequence of almost complex structures $\{J_{\mu,c,R_i}\}_{i=1}^N$ starting at $J_{\mu,c,0}$. We have the holomorphic curve u_0 at this starting point.

To numerically approximate the holomorphic curve u_{R_i} for the almost complex structure J_{μ,c,R_i}, we define the Cauchy–Riemann energy as

$$E_{CR}(u) = \int_{\mathbb{C}P^1} \|du^{0,1}\|^2 dA.$$

Clearly, we have $E_{CR}(J_i, u) = 0$ precisely when u is J_i-holomorphic. Fix N points in $\mathbb{C}P^1$, which we denote by $\{z_1, \ldots, z_N\}$, and approximate $E_{CR}(u)$ by the finite sum,

$$E_{CR,\text{approx}}(J_i, u) = \sum_j w_j \|d_{z_j} u^{0,1}\|^2.$$

In other words, $E_{CR,\text{approx}}$ is a function from N copies of $\mathbb{C}P^2$ to $\mathbb{R}$, and $d_{z_j} u^{0,1}$ is a discretized version of the Cauchy–Riemann operator.

In practice, we choose a good coordinate system on $\mathbb{C}P^1$ using stereographic projection and polar coordinates, and then discretize both the radial and angular coordinate to define the discretized Cauchy–Riemann operator. To keep the discussion simple and avoid the exponential map, we note that we can use a single chart for most of the construction. Indeed, by positivity of intersection and the intersection form of $\mathbb{C}P^2$, any holomorphic curve u not equal to $\mathbb{C}P^1_\infty$ with homology class $[u] = [\mathbb{C}P^1]$ must intersect the line $\mathbb{C}P^1_\infty$ in precisely one point, which we called p. Hence $u|_{\mathbb{C}P^1\setminus\{\infty\}}$ takes values in $\mathbb{C}P^2 - \mathbb{C}P^1_\infty$, which is diffeomorphic to $\mathbb{C}^2$. We may, in this chart, define the discretized Cauchy–Riemann operator by

$$d_{z_{s,t}} u^{0,1} := \frac{u(z_{s+1,t}) - u(z_{s,t})}{z_{s+1,t} - z_{s,t}} + J_{\mu,c,R_i}(u(z_{s,t})) \cdot \frac{u(z_{s,t+1}) - u(z_{s,t})}{z_{s,t+1} - z_{s,t}},$$

where we have chosen a particular collection of points for $\mathbb{C}P^1$ coming from cylindrical coordinates with s corresponding to the radial coordinate and t with the angular coordinate.

To obtain u_{R_i} from $u_{R_{i-1}}$, the idea is to apply the gradient flow to this function $E_{CR,\text{approx}}$ to find minima of $E_{CR,\text{approx}}$, which is a function from $(\mathbb{C}^2)^N \to \mathbb{R}$.

18.3.2 Some implementation details

To speed up the process, we used the following:

1. Computing the gradient of $E_{CR,\text{approx}}$ by varying each value $u(z_{s,t})$ is a local operation. This means that we don't need to recompute the entire sum $E_{CR,\text{approx}}$. Only the nearest neighbors of $z_{s,t}$ are needed.
2. The gradient flow seems to get stuck often in local extrema or points that are nearly local extrema. We apply random perturbations to deal with this and to further decrease the approximate Cauchy–Riemann energy.

The upshot is rather disappointing. This method works well for small perturbations of J_0, but works poorly when we do not approximately know the asymptotics of the desired map u. Adding more resolution by increasing N may help, but Formula (18.3) is much more efficient for constructing a practical foliation that is transverse to the flow as we will see in the following.

18.3.3 An ad hoc approach

Although this method makes sense from the point of view of Poincaré's and Birkhoff's results, it specifically relies on the existence of a pair of orbits that form a Hopf link: we give a conjectural formula based on this. As far as we know, such a claim has not been proved in the literature for all mass ratios and all Jacobi energies below the first critical value.

Fix the mass ratio μ and energy level c, and construct parametrizations of the retrograde orbit γ_r and a direct orbit γ_d by numerical shooting. If γ_d is not unique, we select the one for which $\gamma_d(0)$ has minimal x-coordinate.

Since $\Sigma_{\mu,c}$ is starshaped, there is a well-defined smooth projection $p_{\mu,c}$: $\mathbb{C}^2 \setminus \{0\} \to \Sigma_{\mu,c}$.

To get an open book with disk-pages, we define

$$\begin{aligned} f_D : S^1 \times D^2 &\longrightarrow \Sigma_{\mu,c} \\ (\theta; r, \phi) &\longmapsto p_{\mu,c}\left(r\gamma_r\left(\frac{T_r\phi}{2\pi}\right) + \sqrt{1-r^2}\gamma_d\left(\frac{T_d\theta}{2\pi}\right)\right). \end{aligned} \tag{18.3}$$

For the open book with annulus-pages, we define

$$\begin{aligned} f_A : S^1 \times I \times S^1 &\longrightarrow \Sigma_{\mu,c} \\ (\theta; r, \phi) &\longmapsto p_{\mu,c}\left(r\gamma_r\left(\frac{T_r\phi}{2\pi}\right) + (1-r)\gamma_d\left(\frac{T_d(\theta-\phi)}{2\pi}\right)\right). \end{aligned} \tag{18.4}$$

Note that these are only conjectural formulas: as stated, it is not even clear that these are defined. For instance,

1. The existence of a suitable version of the direct orbit has not been proved without computer-assisted arguments.
2. The argument of the projection, a linear combination, may go through 0, in which case we cannot apply the projection map $p_{\mu,c}$
3. The resulting map, even when it is defined, may not be an embedding away from the binding.

On the positive side, there are some reasons to believe that this should work.

1. The first formula reproduces McGehee's disk-like surface of section, and essentially corresponds to the Poincaré's surface of section in the case of an an-

nulus. Note that Poincaré actually used a half-plane, see Section 305 in Chapter XXVII of Poincaré's *Les Méthodes Nouvelles de la Mécanique Céleste III*, [205].

2. We can check numerically whether this works. We will discuss this below and in Section 18.4.

The intuitive reason that this approach seems to work is that even for large mass ratios and Jacobi energies close to the critical value, the retrograde and direct orbit appear, from the point of view of numerics, to be sufficiently similar to Hopf fibers.

Numerical construction of $p_{\mu,c}$

Since $\Sigma_{\mu,c}$ is starshaped, we know that for a given point $x \in \mathbb{C}^2 \setminus \{0\}$, there is a real number $\lambda(x) > 0$ such that $\lambda(x) \cdot x \in \Sigma_{\mu,c}$. In other words, we are looking for zeros λ_0 of the function $H(\lambda x) + c$. We will find these zeros by Newton iterations, but because $H^{-1}(-c)$ has several connected components, we need to take some care to ensure that the zeros we find actually lie on $\Sigma_{\mu,c}$.

To deal with this, we homotope the Hamiltonian to a simpler Hamiltonian for which $H^{-1}(-c)$ is connected. Then, for each homotopy parameter s we find λ_s using Newton iterations such that $\lambda_s x \in H_s^{-1}(-c)$. Carrying out the homotopy gives us the desired function $p_{\mu,c}$.

Checking the conjectural formulas

We can numerically verify whether the conjectural formulas make sense on the discretized surface S_d as follows.

1. Each pair of points $\gamma_{\mathrm{retro}}(\phi)$ and $\gamma_{\mathrm{direct}}(\theta)$ on the retrograde and direct orbit, respectively, are linearly independent. Interval arithmetic can be used to make this precise for particular ranges of parameter values. It would be interesting to have an analytic proof, but this first requires an analytic proof of the existence of the direct orbit. A corollary is that the maps f_D and f_A are defined, i.e., the argument of the projection $p_{\mu,c}$ never vanishes.
2. At the discretized surface S_d, the Hamiltonian vector field can be seen to be positively transverse by showing that the tangent vector fields to the disk, the Hamiltonian vector field and the Liouville vector field form a basis of $\mathbb{C}^2$. Of course this breaks down at the boundary of the surface, and this complicates a possible numerical proof with interval arithmetic: we have no idea how to give a proof.

We also mention at this point Conley's construction of a global surface of section in [57], which is essentially the same annulus as the one above: he used power series expansions to find the retrograde and direct orbit. These series converge if c is sufficiently large.

18.4 Finding a discretized return map and seeing the dynamics

We consider the leaves of a foliation by global surfaces of section $\{L_\theta\}$ for $\theta \in [0, 2\pi]$, where each leaf is diffeomorphic to a compact surface S, so we have the map

$$f : S \times S^1 \to \Sigma_{\mu,c}$$

with $L_\theta = f(S, \theta)$. For example, we can set $L_\theta = f_D(D^2, \theta)$.

We will find a numerical approximation of the return map $\psi : L_0 \to L_0$. For this, fix a finite set S_d of distinct points in the interior of S. In practice, we apply the formulas (18.3) and (18.4) directly to finite sets S_d discretizing the disk and annulus, respectively. Choose a map $p_d : S \to S_d$, which sends a point x to the closest point in S_d: the choice of the closest point can be non-unique, although in numerical practice this is not an issue. Define the discretized return map as $\psi_d := p_d \circ \psi|_{S_d} : S_d \to S_d$. Although we have seen in Proposition 17.1.9 that $\psi|_S$ is conjugated to an area-preserving diffeomorphism of $int(S)$, the map ψ_d is usually not a bijection.

Since the surfaces of section are now given by some complicated, non-linear equation, it is hard to detect when the flow of a point has crossed the surface L_0. To deal with this, we compute the return map in several steps, see the sketch in Figure 18.7.

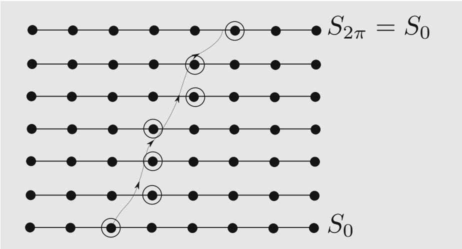


Figure 18.7: Using the foliation to obtain the discretized return map.

1. Take a finite collection of leaves $\{L_{k\epsilon}\}_{k=0}^{N}$, such that $L_{N\epsilon} = L_{2\pi}$. We clarify the choice of ϵ below
2. Take a point p in L_0: flow until we hit the next leaf L_ϵ. If ϵ is sufficiently small, then the flowed point will be close to $f(p, \epsilon)$.
3. Reaching the last leaf requires extra care.

For the choice of ϵ, consider an orbit γ intersecting the leaf L_0 in the point p_0 and intersecting the leaf L_ϵ in p_ϵ. Then we want to require that the preimages in the abstract disk, which are given by $z_0 = f^{-1}(\cdot, 0)(p_0)$ and $z_\epsilon = f^{-1}(\cdot, \epsilon)(p_\epsilon)$, are sufficiently close. Since the Reeb vector field is transverse to the interior of leaves, it is possible to make this choice. However, a uniform choice is not possible since transversality fails at the boundary. This is fixed by some experimentation with the parameter ϵ. Outside a tubular neighborhood of the binding orbits, we can compute a good value of ϵ by looking at the derivative of the open book projection map to S^1 with respect to the Hamiltonian vector field. Closer to the binding, we perform a more careful search for the closest point.

18.4.1 Some results and observations

Annulus maps

We start by applying this method to annulus-type surfaces of section using Formula (18.4) to construct the foliation by annuli. We will describe the return map by seeing where a lattice in $\mathbb{R}/2\pi\mathbb{Z} \times (0,1)$ maps. We use coordinates $\phi \in \mathbb{R}/2\pi\mathbb{Z}$ and $r \in (0,1)$.

Traditionally, the so-called monotone twist maps are of particular importance due to Poincaré's last geometric theorem. We first give a general definition which does not seem to be in common use. We will call a symplectomorphism ψ of a Liouville cobordism (W, ω, X) a *monotone twist map* if $(\psi^*\lambda)(X) \neq 0$. Here $\lambda = i_X\omega$ is the Liouville form associated with the Liouville vector field X. The most common situation is when $W = (0,1) \times S^1$ with symplectic form $dr \wedge d\phi$ and Liouville vector field $X = r\partial_r$. We may write ψ in components, $\psi(r_0, \phi_0) = (\psi_r(r_0, \phi_0), \psi_\phi(r_0, \phi_0))$. In this case $\lambda = d\phi$ and the monotone twist condition just means that

$$d\phi(d\psi r\partial_r) = d\phi(r\partial_r\psi) = r \cdot \frac{\partial \psi_\phi}{\partial r} \neq 0,$$

which means that vertical rays $(0,1) \times \{\phi_0\}$ are mapped to curves that bend either left or right, but do not go straight at any point.

Roughly speaking, Poincaré already observed that the first return map of the planar restricted three-body problem is a monotone twist map for small μ, and Jacobi energy c below the first Lagrange value.

Since we can now easily construct a (conjectural) foliation by annuli with Formula (18.4), we compute the return map for other mass ratios. In the following two figures, it is clear that the twist condition is violated in the (ϕ, r)-coordinates. This does not prove that the twist-condition does not hold; in different coordinates it may hold. We also point out the maps are area-preserving for the area form induced by $d\lambda$. This area form differs from the standard area form, meaning that the lattice does not reflect the area directly.

Graphs of disk maps

The main topic of this monograph is disk-like surfaces of section and we will now see what dynamics we can find in the restricted three-body problem using such

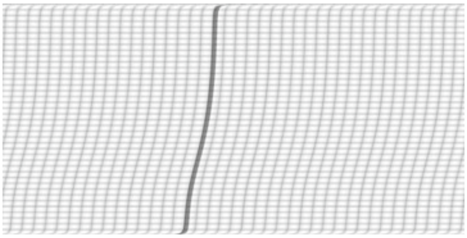


Figure 18.8: Return map for $\mu = 0$ and $c = 2.0$: the rectangular lattice is deformed by an integrable twist. A vertical segment (dark) is mapped to the segment bending to the right.

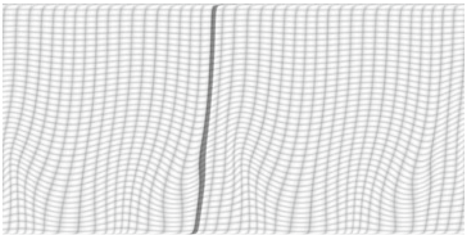


Figure 18.9: Return map for $\mu = 0.23$ and $c = 2.0$: the rectangular lattice is deformed.

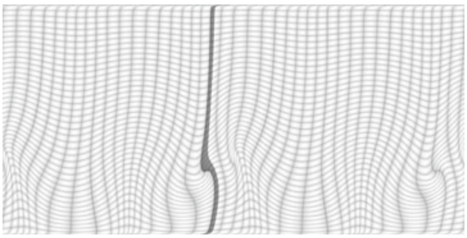


Figure 18.10: Return map for $\mu = 0.5325$ and $c = 2.0$: the rectangular lattice is strongly deformed. A vertical segment (dark) is mapped to the curve bending both left and right.

an object. We will draw graphs of the return maps $\psi^k : D^2 \to D^2$ by color-coding the points: given $(x, y) \in D^2$, we color the point by giving it $\frac{x+1}{2}$-redness and $\frac{y+1}{2}$-greenness in an red-green-blue-decomposition.

Note that the accuracy of this method is rather limited. Although we can ensure that the first return map is accurate up to the discretization error of the surface S_d, this error will quickly propagate making high iterates rather unreliable.

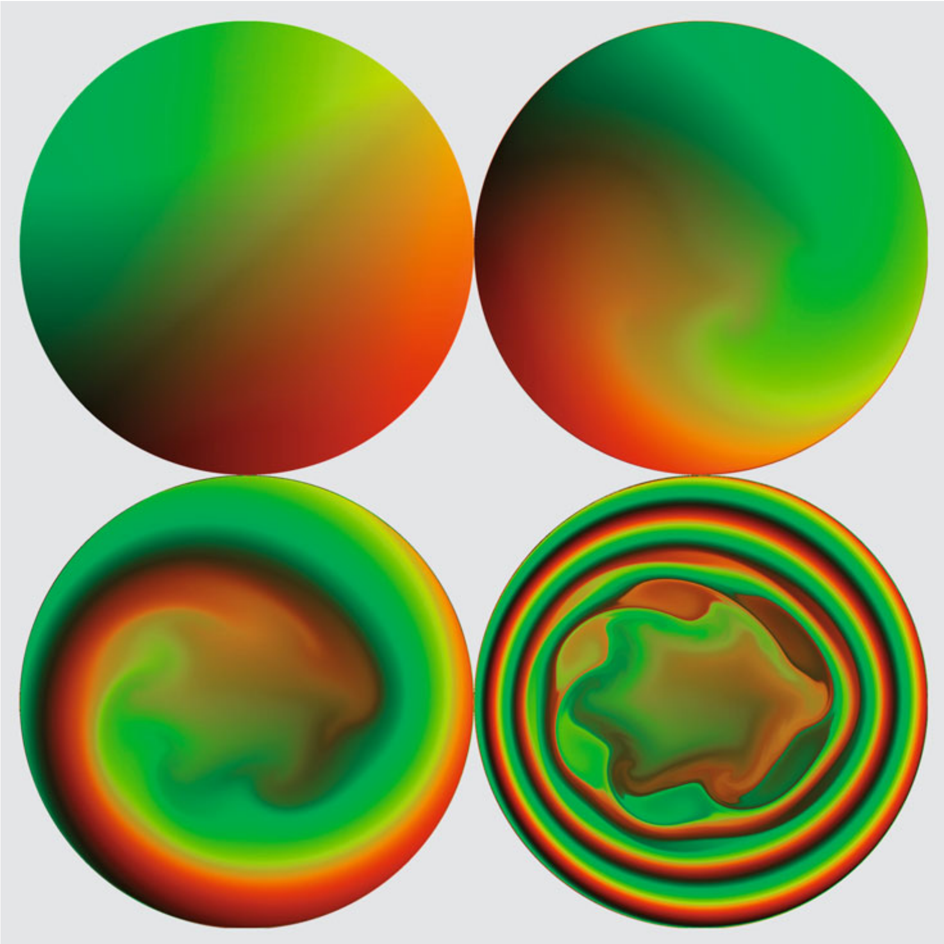


Figure 18.11: Return maps for $\mu = 0.48$ and $c = 2.0$: the identity, the first return map, a 5-fold iterate, and a 20-fold iterate. Some elliptic islands are clearly visible; in between there are some hyperbolic periodic points.

This seems not to be a big problem for small mass ratios, where the system is nearly integrable, but quickly becomes severe for larger mass ratios.

Collision orbits

Obviously, those points whose orbits undergo a two-body collision, i.e., the q-coordinate becomes 0 in the unregularized problem, form an invariant set $C_{\mu,c}$, defined by

$$C_{\mu,c} = \{p \in i(D) \mid \text{there is a sequence } \{t_n\}_n \text{ such that } \lim_{n\to\infty} Q(\phi^{t_n}_{X_H}(p)) = 0\}.$$

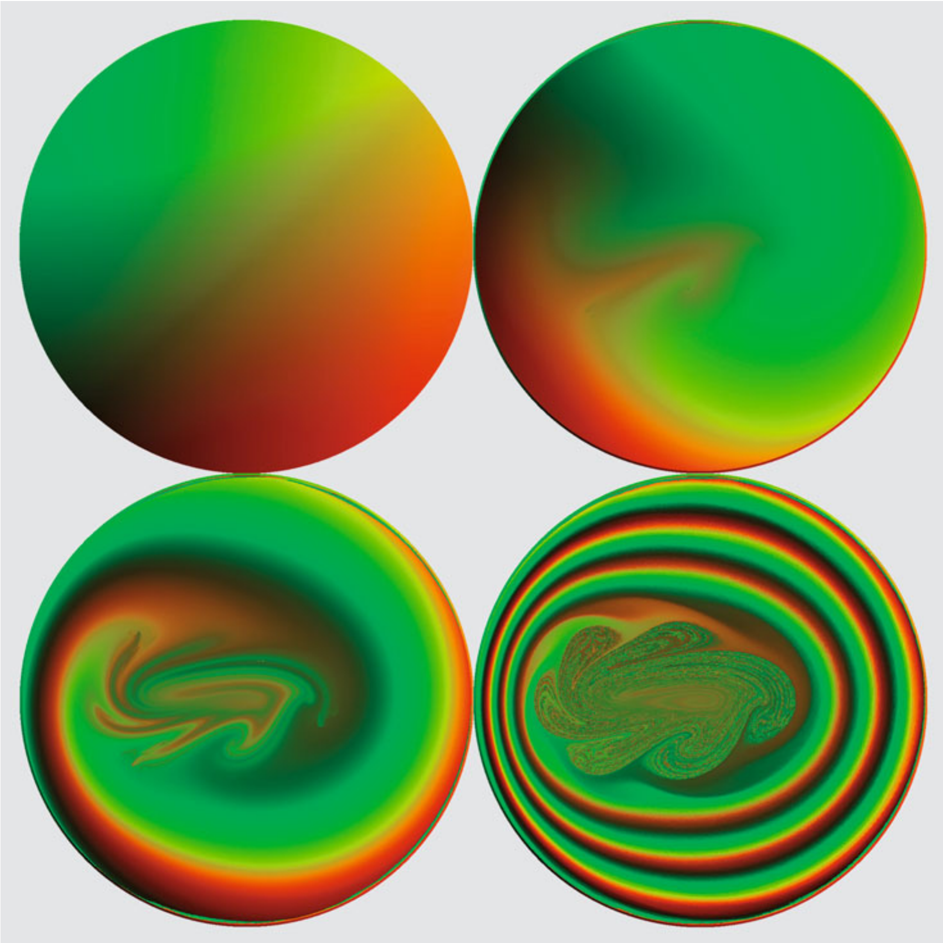
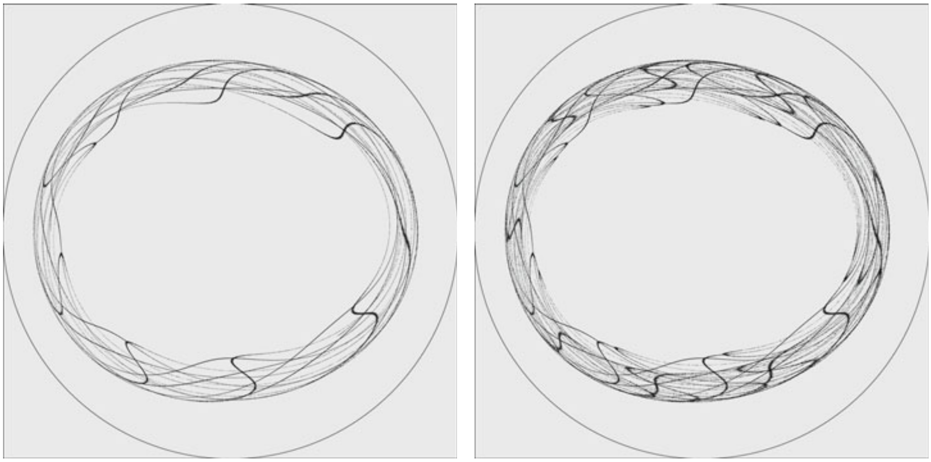


Figure 18.12: Return maps for the Moon-Earth very close to the first critical value $\mu = 0.9878$ and $c = 1.59417$: the identity, the first return map, a 5-fold iterate, and a 20-fold iterate. There is a clear elliptic orbit in the center, but the surrounding area is almost completely stochastic, and hence our approximation is not reliable. Note the not-so-surprising similarity with the work of Simó and Stuchi on the Hill's lunar problem, [222], who used the traditional (non-global) surface of section. The projection of our global surface of section to the (q_1, q_2)-plane (not shown here) explains why this traditional approach has worked well near the center.

We will call this the *collisional set*[6] to distinguish it from the collision set, which is not invariant under the flow.

[6]We thank Lei Zhao for coining this term.

For $\mu = 0$, we see directly from McGehee's description that the collisional set $C_{0,c}$ forms an invariant curve. For positive μ, the collisional set can be more interesting as we see in Figure 18.13. This leads us to the following question.



(a) Collision within five iterates. (b) Collision within ten iterates.

Figure 18.13: The curve indicates those points on a global surface of section that undergo a collision within the first five iterates (left) and within the first ten iterates (right). Both figures are for mass ratio $\mu = 0.9$ near the light primary, and for Jacobi energy close to the first Lagrange value. These curves form a subset of the collisional set.

Question: Can the closure of the collisional set $C_{\mu,c}$ in the surface of section have positive measure?

Computing more iterates numerically suggests (but does not prove) that the answer is yes for $\mu > 0$.

18.4.2 Another return map

The return map in Figure 18.14 is for a mass ratio that is close to that of the Pluto-Charon system. We are looking at a Jacobi energy that is very close (but below) the first critical value.

The picture on the lower right is very reminiscent of Arnold's famous drawing: we see approximate KAM tori near the boundary, because the boundary is the retrograde orbit, which is here elliptic. We see the thin elliptic islands and hyperbolic periodic points. Finally the homoclinic chaos is also very apparent.

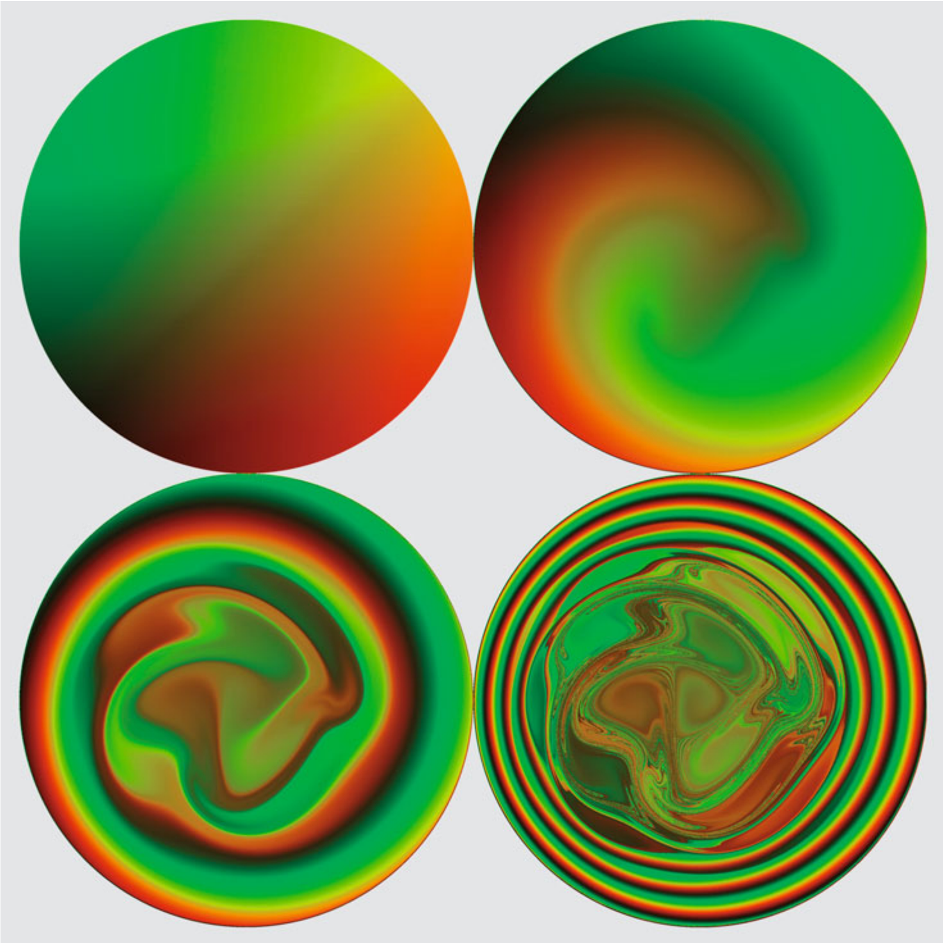


Figure 18.14: Return maps for $\mu = 0.1$ and c very close to L_1: the identity, the first return map, a 5-fold iterate, and a 20-fold iterate.

Bibliography

[1] C. Abbas, H. Hofer, *Holomorphic Curves and Global Questions in Contact Geometry*, book in preparation.

[2] A. Abbondandolo, B. Bramham, U. Hryniewicz, P. Salomão, *Sharp systolic inequalities for Reeb flows on the three-sphere*, arXiv:1504.05258

[3] R. Abraham, J. Marsden, *Foundations of Mechanics*, 2nd ed., Addison-Wesley, Reading (1978).

[4] R. Abraham, J. Marsden, T. Ratiu, *Manifolds, tensor analysis, and applications*, 2nd ed., Applied Mathematical Sciences **75**, Springer (1988).

[5] B. Aebischer, M. Borer, M. Kalin, C. Leuenberger, *Symplectic Geometry*, Progress in Mathematics **124**, Birkhäuser, Basel (1994).

[6] P. Albers, J. Fish, U. Frauenfelder, H. Hofer, O. van Koert, *Global surfaces of section in the planar restricted* 3*-body problem*, Arch. Ration. Mech. Anal. **204** (2012), no. 1, 273–284.

[7] P. Albers, J. Fish, U. Frauenfelder, O. van Koert, *The Conley–Zehnder indices of the rotating Kepler problem*, Math. Proc. Cambridge Phil. Soc. **154**, no. 2 (2013), 243–260.

[8] P. Albers, U. Frauenfelder, *Rabinowitz Floer homology: A survey*, in Global Differential Geometry, Springer Proc. in Math. (2012), 437–461.

[9] P. Albers, U. Frauenfelder, O. van Koert, G. Paternain, *Contact geometry of the restricted three-body problem*, Comm. Pure Appl. Math. **65** (2012), no. 2, 229–263.

[10] P. Albers, H. Hofer, *On the Weinstein conjecture in higher dimensions*, Comm. Math. Helv. (2009), 429–436.

[11] A. Albouy, *There is a projective dynamics,* EMS Newsletter **89** (2013), 37–43.

[12] J. Alexander, *A lemma on systems of knotted curves*, Proc. Nat. Acad. Sci. U.S.A. 9 (1923), 93–95.

[13] J. Álvarez Paiva, F. Balacheff, *Contact geometry and isosystolic inequalities*, Geom. Funct. Anal. **24** (2014), 648–669.

[14] P. Appell, *De l'homographie en mécanique,* American Journal of Mathematics, **12** (1890), 103–114.

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*, Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8

[15] R. Arenstorf, *Periodic solutions of the restricted tree-body problem representing analytic continuations of Keplerian elliptic motions*, Am. J. Math. **85** (1963), 27–35.

[16] V. Arnold, *On a characteristic class entering into conditions of quantization*, Func. Anal. **1** (1967), 1–8.

[17] V. Arnold, *Dynamical Systems V: Bifurcation Theory and Catastrophe Theory*, Encyclopaedia of Mathematical Sciences **5**, Springer Berlin Heidelberg (1994).

[18] V. Arnold, *Mathematical methods of classical mechanics*, Graduate texts in mathematics **60**, New York, Springer (2000).

[19] V. Arnold, A. Givental, *Symplectic topology*, In *Dynamical Systems* **IV**. Springer-Verlag, Berlin (1990).

[20] V. Arnold, V. Kozlov, A. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, 3. edition, Springer-Verlag Berlin Heidelberg (2006).

[21] M. Atiyah, *Convexity and commuting Hamiltonians*, Bull. London Math. Soc. **14** (1982), 1–15.

[22] H. Bacry, H. Ruegg, J. Souriau, *Dynamical Groups and Spherical Potentials in Classical Mechanics*, Commun. math. Phys. **3** (1966), 323–333.

[23] V. Bangert, Y. Long, *The existence of two closed geodesics on every Finsler 2-sphere*, Math. Ann. **346** (2010), 335–366.

[24] V. Bargmann, *Zur Theorie des Wasserstoffatoms: Bemerkungen zur gleichnamigen Arbeit von V. Fock*, Zeitschrift für Physik **99** (1936), 576–582.

[25] R. Barrar, *Existence of periodic orbits of the second kind in the restricted problem of three bodies* Astron. J. **70** (1965), 3–5.

[26] R. Barrar, *Periodic orbits of the second kind*, Indiana Univ. Math. J. **22** (1972), 33–41.

[27] J. Barrow-Green, *Poincaré and the Three Body Problem*, AMS-LMS (1997).

[28] R. Bate, D. Mueller, J. White, *Fundamentals of Astrodynamics*, Dover Publications; Revised ed. edition (1971)

[29] A. Bathkin, N. Bathkina, *Hénon's generating solutions and the structure of periodic orbits families of the restricted three-body problem*, arXiv:1411.4933

[30] E. Belbruno, *Two body motion under the inverse square central force and equivalent geodesic flows*, Celest. Mech. **15** (1977), 467–476.

[31] E. Belbruno, *Regularizations and geodesic flows*, Lecture notes in Pure and Applied Mathematics **80** (1981), 1–11.

[32] E. Belbruno, *Lunar capture orbits, a method of constructing earth-moon trajectories and the lunar gas mission*, Proceedings of AIAA/DGGLR/JSASS Inter. Elec. Propl. Conf., number 87-1054 (1987).

[33] E. Belbruno, *Capture Dynamics and Chaotic Motions in Celestial Mechanics*, Princeton University Press (2004).

[34] E. Belbruno, *Fly me to the moon*, Princeton University Press: Princeton (2007).

[35] E. Belbruno, J. Miller, *A ballistic lunar capture trajectory for the Japanese spacecraft Hiten*, Technical Report JPL-IOM 312/90.4-1731-EAB, Jet Propulsion Laboratory (1990).

[36] J. van den Berg, F. Pasquotto, C. Vandervorst, *Closed characteristics on non-compact hypersurfaces in* $\mathbb{R}^{2n}$, Math. Ann. **343** (2009), 247–284.

[37] M. Berger, *A Panoramic View of Riemannian Geometry*, Springer-Verlag Berlin Heidelberg (2003).

[38] G. Birkhoff, *Proof of Poincaré's geometric theorem*, Trans. Amer. Math. Soc. **14** (1913), 14–22.

[39] G. Birkhoff, *The restricted problem of three bodies*, Rend. Circ. Matem. Palermo **39** (1915), 265–334.

[40] A. Bolsinov, A. Fomenko, *Integrable Hamiltonian systems. Geometry, topology, classification*, translated from the 1999 Russian original. Chapman & Hall/CRC, Boca Raton, FL, 2004. xvi+730 pp. ISBN: 0-415-29805-9

[41] R. Bott, L. Tu, *Differential forms in algebraic topology*, Graduate Texts in Mathematics, 82. Springer-Verlag, New York-Berlin, 1982. xiv+331 pp. ISBN: 0-387-90613-4

[42] F. Bourgeois, *A Morse–Bott approach to contact homology*, Ph.D. thesis, Stanford University (2002).

[43] F. Bourgeois, Y. Eliashberg, H. Hofer, K. Wysocki, E. Zehnder, *Compactness results in Symplectic Field theory*, Geom. and Top. **7** (2003), 799–889.

[44] T. Bröcker, K. Jänich, *Introduction to Differential Topology*, Cambridge University Press (1982).

[45] R. Broucke, *Periodic orbits in the restricted three-body theorem with Earth-Moon masses*, Pasadena, Jet Propulsion Laboratory, California Institute of Technology (1968).

[46] L. Brouwer, *Beweis des ebenen Translationssatzes*, Math. Ann. **72** (1912), 37–54.

[47] L. Brouwer, *Über die periodischen Transformationen der Kugel*, Math. Ann. **80** (1919), 39–41.

[48] E. Brown, *An Introductory Treatise on the lunar Theory*, Cambridge University Press (1896).

[49] A. Bruno, *The restricted* 3*-body problem: Plane periodic orbits*, De Gruyter Expositions in Mathematics **17**, Walter de Gruyter, Berlin, New York (1994).

[50] A. Cannas da Silva, *Lectures on Symplectic geometry*, 2nd printing, Lecture Notes in Mathematics **1764**, Springer (2008).

[51] C. Charlier, *Die Mechanik des Himmels*, Leipzig, Veit (1902).

[52] A. Chenciner, R. Montgomery, *A remarkable periodic solution of the three-body problem in the case of equal masses*, Ann. of Math. (2) 152 (2000), no. 3, 881–901.

[53] A. Chenciner (2007), *Three body problem*, Scholarpedia, 2(10):2111.

[54] A. Chenciner, *Poincaré and the Three-Body Problem*, Poincaré 1912–2012, Séminaire Poincaré XVI (2012), 45–133.

[55] K. Cieliebak, U. Frauenfelder, *A Floer homology for exact contact embeddings*, Pacific J. Math. 239 (2009), no. 2, 251–316.

[56] K. Cieliebak, U. Frauenfelder, O. van Koert, *The Finsler geometry of the rotating Kepler problem.* Publ. Math. Debrecen 84 (2014), no. 3-4, 333–350.

[57] C. Conley, *On Some New Long Periodic Solutions of the Plane Restricted Three Body Problem*, Comm. Pure Appl. Math. **16** (1963), 449–467.

[58] C. Conley, *The Retrograde Circular Solutions of the Restricted Three-Body Problem Via a Submanifold Convex to the Flow*, SIAM J. Appl. Math. **16**, no. 3, (1968), 620–625.

[59] C. Conley, *Low energy transit orbits in the restricted three-body problem*, SIAM J. Appl. Math. **16** (1968), 732–746.

[60] C. Conley, *Twist mappings, linking, analyticity, and periodic solutions which pass close to an unstable periodic solution*, Topological dynamics (Sympos., Colorado State Univ., Ft. Collins, Colo., 1967), Benjamin, New York (1968), 129–153.

[61] C. Conley, *On the ultimate behavior of orbits with respect to an unstable critical point I, oscillating, asymptotic and capture orbits*, J. Differential Equations **5** (1969), 136–158.

[62] C. Conley, E. Zehnder, *The Birkhoff–Lewis fixed point theorem and a conjecture of V.I. Arnold,* Invent. Math. **73** (1983), 33–49.

[63] C. Conley–Zehnder, *Morse-type index theory for flows and periodic solutions of Hamiltonian equations,* Commun. Pure Appl. Math. **37** (1984), 207–253.

[64] A. Constantin, B. Kolev, *The theorem of Kérékjartò on periodic homeomorphisms of the disc and the sphere*, Enseign. Math. (2) **40** (1994), no. 3–4, 193–204.

[65] D. Cristofaro-Gardiner, M. Hutchings, *From one Reeb orbit to two*, arXiv: 1202.4839

[66] R. Cushman, L. Bates, *Global aspects of classical integrable systems* Second edition. Birkhäuser/Springer, Basel, 2015. xx+477 pp. ISBN: 978-3-0348-0917-7

[67] R. Cushman, J. Duistermaat, *A Characterization of the Ligon–Schaaf Regularization Map*, Comm. Pure Appl. Math. **50** (1997), 773–787.

[68] G. Darwin, *Periodic orbits*, Acta Math. **21**, (1897), 99–242.

[69] G. Darwin, *On certain families of periodic orbits*, Monthly Notices Roy. Astron. Soc. **70**, (1909). 108–143.

[70] C. Delaunay, *Nouvelle théorie analytique du mouvement de la lune*, Comptes Rendus **23**, 968–970.

[71] C. Delaunay, *Théorie du mouvement de la lune I*, Mémoire de l'Academie des Sciences **28**, 1–883.

[72] C. Delaunay, *Théorie du mouvement de la lune II*, Mémoire de l'Academie des Sciences **29**, 1–931.

[73] T. Delzant, *Hamiltoniens périodiques et image convex de l'application moment*, Bull. Soc. Math. France **116** (1988), 315–339.

[74] L. Doyle, W. Welsh, *Worlds with Two Suns*, Scientific American **309**(5), (2013), 40–47.

[75] J. Duistermaat, *On the Morse index in Variational Calculus*, Adv. Math. **21** (1976), 173–195.

[76] C. Ehresmann, *Les connexions infinitésimales dans un espace fibré différentiable*, Colloque de Topologie, Bruxelles (1950), 29–55.

[77] Y. Eliashberg, A. Givental, H. Hofer, *Introduction to symplectic field theory*, Geom. Funct. Anal. **10** (2000), 560–673.

[78] L. Euler, *Un corps étant attiré en raison déciproque quarrée des distances vers deux points fixes donnés*, Mémoires de l'Acad. de Berlin (1760), 228–249.

[79] L. Euler, *De motu corporis ad duo centra virium fixa attracti*, Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae **10** (1766), 207–242, **11** (1767), 152–184.

[80] R. Farquhar, *The control and use of libration-point satellites*, Technical Report TR R-346, NASA (1970).

[81] R. Farquhar, D. Muhonen, C. Newman, H. Heuberger, *Trajectories and orbital maneuvers for the first libration-point satellite*, J. Guid. and Control **3** (1980), 549–554.

[82] A. Fathi, *An orbit closing proof of Brouwer's lemma on translation arcs*, L'enseignement Math. **33** (1987), 315–322.

[83] S. Ferraz-Mello (2009), *Celestial mechanics*, Scholarpedia, 4(1):4416.

[84] J. Fish, R. Siefring, *Connected sums and finite energy foliations I: Contact connected sums*, arXiv:1311.4221

[85] A. Floer, *Morse theory for Lagrangian intersections*, J. Differential Geom. **28** (1988), 513–547.

[86] A. Floer, H. Hofer, D. Salamon, *Transversality in elliptic Morse theory for the symplectic action*, Duke Math. Journal, **80** (1996), 251–292.

[87] V. Fock, *Zur Theorie des Wasserstoffatoms*, Zeitschrift für Physik **98** (1935), 145–154.

[88] A. Fomenko, *Symplectic geometry*, Advanced studies in contemporary mathematics **5**, New York, Gordon and Breach (1988).

[89] J. Franks, *A new proof of the Brouwer plane translation theorem*, Erg. Th. and Dyn. Systs., **12** (1992), 217–226.

[90] J. Franks, *Geodesics on S^2 and periodic points of annulus homeomorphisms*, Invent. Math., **108** (1992), 403–418.

[91] J. Franks, *Area preserving homeomorphisms of open sufaces of genus zero*, New York Jour. Math. **2** (1996), 1–19.

[92] U. Frauenfelder, *Dihedral homology and the moon*, J. Fixed Point Theory Appl. **14** (2013), 55–69.

[93] U. Frauenfelder, J. Kang, *Real holomorphic curves and invariant global surfaces of section*, Proc. Lond. Math. Soc. (3) 112 (2016), no. 3, 477–511.

[94] U. Frauenfelder, O. van Koert, L. Zhao, *A convex embedding for the rotating Kepler problem*, arXiv:1605.06981

[95] D. Fuks, *The Maslov–Arnold characteristic classes*, Dokl. Akad. Nauk SSSR **178** (1968), 303–306.

[96] N. Gavrilov, A. Shilnikov, *Example of a blue sky catastrophe*, AMS Transl. Series II, v.200, 99–105, 2000.

[97] H. Geiges, *An introduction to contact topology*, Cambridge Studies in Adv. Math. **109**, Cambridge Univ. Press (2008).

[98] H. Geiges, *The Geometry of Celestial Mechanics*, book in progress.

[99] J. Ginsburg, D. Smith, *A History of Mathematics in America before* 1900, The Mathematical Association of America (Carus Mathematical Monograph Number 5).

[100] V. Ginzburg, V. Guillemin, Y. Karshon, *Moment Maps, Cobordism, and Hamiltonian Group Actions*, Math. Surv. and Monographs, **98**. Amer. Math. Soc.

[101] V. Ginzburg, B. Gürel, *A C^2-smooth counterexample to the Hamiltonian Seifert conjecture in $\mathbb{R}^4$*, Ann. of Math. (2) **158** (2003), no.3, 953–976.

[102] H. Goldstein, *Prehistory of the Runge–Lenz vector*, Amer. Jour. Phys. **43** (1975), no. 8, 737–738.

[103] H. Goldstein, *More on the prehistory of the Runge–Lenz vector*, Amer. Jour. Phys. **44** (1976), no. 11, 1123–1124.

[104] E. Goursat, *Sur les transformations isogonales en Mécanique*, Comptes Rendus des Séances de l'Académie des Sciences, Paris, **108**, (1889), 446–448.

[105] M. Gromov, *Pseudo Holomorphic curves in symplectic manifolds*, Invent. Math. **82** (1985), 307–347.

[106] V. Guillemin, S. Sternberg, *Convexity Properties of the Moment Mapping*, Inv. Math. **67** (1982), 491–513.

[107] L. Guillou, *Théorème de translation plane de Brouwer et généralisations du théorème de Poincaré–Birkhoff*, Topology **33**(2) (1994), 331–351.

[108] G. Györgyi, *Kepler's equation, Fock variables, Bacry's generators and Dirac brackets*, Nuovo Cimento, 53A (1968), 717–735.

[109] Y. Hagihara, *Celestial mechanics*, **1-2**, MIT Press, Cambridge, Massachusetts (1970–1972); **3-5**, Japan Society for the Promotion of Science, Tokyo, (1974–1976).

[110] G. Hall, K. Meyer, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Applied Mathematical Sciences **90**, Springer (1992).

[111] A. Hatcher, *Algebraic Topology*, Cambridge University Press (2001).

[112] G. Heckman, T. de Laat, *On the Regularization of the Kepler Problem*, Jour. Symp. Geom. **10**, vol. 3 (2012), 463–473.

[113] G. Heckman, L. Zhao, *Angular Momenta of Relative Equilibrium Motions and Real Moment Map Geometry*, Invent. Math. 205 (2016), no. 3, 671–691.

[114] M. Hénon, *Numerical exploration of the restricted problem. V. Hill's case: Periodic orbits and their stability*, Astron. Astrophys. **1**, 223–238.

[115] M. Hénon, *Generating Families in the Restricted Three-Body Problem*, Springer (1997).

[116] G. Hill, *On the part of the Motion of the Lunar Perigee which is a Function of the Mean Motions of the Sun and the Moon*, John Wilson & Son, Cambridge, Massachusetts (1877), Reprinted in Acta **8**, (1886), 1–36.

[117] G. Hill, *Researches in the lunar theory*, Amer. J. Math.**1** (1878), 5–26, 129–147, 245–260.

[118] A. Hiltebeitel, *On the problem of two fixed centres and certain of its generalizations*, Amer. J. Math. **33**, no. 1/4 (1911), 337–362.

[119] L. Hörmander, *Fourier integral operators I*, Acta Math. **127** (1971), 79–183.

[120] H. Hofer, *Pseudoholomorphic curves in symplectisations with application to the Weinstein conjecture in dimension three*, Invent. Math. **114** (1993), 515–563.

[121] H. Hofer, K. Wysocki, E. Zehnder, *A characterisation of the tight three-sphere*, Duke Math. J. **81** (1995), 159–226.

[122] H. Hofer, K. Wysocki, E. Zehnder, *Properties of pseudo-holomorphic curves in symplectizations.* II. *Embedding controls and algebraic invariants*, Geom. Funct. Anal. **5** (1995), no. 2, 270–328.

[123] H. Hofer, K. Wysocki, E. Zehnder, *Unknotted periodic orbits for Reeb flows on the three-sphere*, Topol. Methods Nonlinear Anal. **7**, (1996), no. 2, 219–244.

[124] H. Hofer, K. Wysocki, E. Zehnder, *Properties of pseudo-holomorphic curves in symplectizations.* I. *Asymptotics*, Ann. Inst. H. Poincaré Anal. Non Linéaire **13** (1996), no. 3, 337–379.

[125] H. Hofer, K. Wysocki, E. Zehnder, *Correction to: "A characterisation of the tight three sphere"*, Duke Math. J. **89** (1997), no. 3, 603–617.

[126] H. Hofer, K. Wysocki, E. Zehnder, *The dynamics on a strictly convex energy surface in* $\mathbb{R}^4$, Ann. Math., **148** (1998) 197–289.

[127] H. Hofer, K. Wysocki, E. Zehnder, *A characterization of the tight* 3*-sphere. II*, Comm. Pure Appl. Math. **52**, no. 9 (1999), 1139–1177.

[128] H. Hofer, K. Wysocki, E. Zehnder, *Properties of pseudo-holomorphic curves in symplectizations.* III*. Fredholm theory*, Topics in nonlinear analysis, Progr. Nonlinear Differential Equations Appl., **35**, Birkhäuser, Basel (1999), 381–475.

[129] H. Hofer, K. Wysocki, E. Zehnder, *Pseudoholomorphic curves and dynamics in three dimensions*, Handbook of dynamical systems, Vol. 1A, North-Holland, Amsterdam (2002), 1129–1188.

[130] H. Hofer, K. Wysocki, E. Zehnder, *Finite energy foliations of tight three-spheres and Hamiltonian dynamics*, Ann. Math. **157** (2003), 125–257.

[131] H. Hofer, E. Zehnder, *Symplectic invariants and Hamiltonian dynamics*, Reprint of the 1994 edition. Modern Birkhäuser Classics. Birkhäuser Verlag, Basel, 2011. xiv+341 pp. ISBN: 978-3-0348-0103-4

[132] W. Hohmann, *Die Erreichbarkeit der Himmelskörper*, Oldenbourg, Munich (1925).

[133] U. Hryniewicz, *Fast finite-energy planes in symplectizations and applications*, Trans. Amer. Math. Soc. **364** (2012), 1859–1931.

[134] U. Hryniewicz, *Systems of global surfaces of section for dynamically convex Reeb flows on the* 3*-sphere*, J. Symplectic Geom. 12 (2014), no. 4, 791–862.

[135] U. Hryniewicz, P. Salomão, *On the existence of disk-like global sections for Reeb flows on the tight* 3*-sphere*, Duke Math. J. **160**, no. 3, (2011), 415–465.

[136] U. Hryniewicz, P. Salomão, *Elliptic bindings for dynamically convex Reeb flows on the real projective three-space*, Calc. Var. Partial Differential Equations 55 (2016), no. 2, Art. 43, 57 pp.

[137] L. Hulthén, *Über die quantenmechanische Herleitung der Balmerterme*, Zeitschrift für Physik **86** (1933), 21–23.

[138] C. Hummel, *Gromov's compactness theorem for pseudo-holomorphic curves*, Progress in Mathematics, 151. Birkhäuser Verlag, Basel, 1997. viii+131 pp. ISBN: 3-7643-5735-5

[139] M. Hutchings, *An index inequality for embedded pseudoholomorphic curves in symplectizations*, J. Eur. Math. Soc. (JEMS) **4** (2002), 313–361.

[140] C. Jacobi, *Vorlesungen über Dynamik*, Berlin, Reimer (1866).

[141] P. Kahn, *Pseudohomology and Homology*, math.AT/0111223.

[142] J. Kang, *Some remarks on symmetric periodic orbits in the restricted three-body problem*, Disc. Cont. Dyn. Sys. (A), **34**, no. 12 (2014), 5229–5245.

[143] J. Kang, *On reversible maps and symmetric periodic points*, arXiv:14103997

[144] T. Kato, *On the convergence of the perturbation method* I, Progr. Theor. Phys. **4** (1949), 514–523.

[145] T. Kato, *On the convergence of the perturbation method* II, Progr. Theor. Phys. **5** (1950), 207–212.

[146] T. Kato, *Perturbation theory for Linear Operators*, Springer, Grundlehren edition (1976).

[147] B. von Kérékjartò, *Über die periodischen Transformationen der Kreisscheibe und der Kugelfläche*, Math. Ann. **80**, (1919–1920), 36–38.

[148] W. Killing, *Die Mechanik in den Nicht-Euklidischen Raumformen*, Jour. reine und angewandte Math. **98** (1885), 1–48.

[149] D. Kim, *Planar Circular Restricted Three Body Problem*, Master thesis, Seoul National University (2011).

[150] S. Kim, *Hamiltonian mechanics and Symmetries*, Master thesis, Seoul National University (2014).

[151] S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry I & II*, Wiley Classics Library (1996).

[152] W. Koon, M. Lo, J. Marsden, S. Ross, *Dynamical Systems, the Three-Body Problem and Space Mission Design*, Marsden Books (2011).

[153] V. Kozlov, A. Harin, *Kepler's problem in constant curvature spaces*, Celestial Mech. Dyn. Astron. **54**, no. 4 (1992), 393–399.

[154] M. Kriener, *An intersection formula for finite energy half cylinders*, PhD thesis, ETH Zurich (1998).

[155] M. Kummer, *On the stability of Hill's Solutions of the Plane Restricted Three Body Problem*, Amer. J. Math. **101**(6), (1979), 1333–1354.

[156] P. Kustaanheimo, E. Stiefel, *Perturbation theory of Kepler motion based on spinor regularization*, J. Reine Angew. Math. **218**,(1965), 204–219.

[157] J. Lagrange, *Recherches sur la mouvement d'un corps qui est attiré vers deux centres fixes*, Miscellania Taurinensia **4** (1766–69), 67–121, Mécanique Analytique, 2nd edition, Paris (1811), 108–121.

[158] J. Lee, *Fiberwise Convexity of Hill's lunar problem*, J. Topol. Anal. 9 (2017), no. 4, 571–630.

[159] J. Lee, *Spectrum estimates of Hill's lunar problem*, arXiv:1510.07248

[160] T. Levi-Civita, *Sur la régularisation du problème des trois corps*, Acta Math. **42** (1920), 99–144.

[161] P. Libermann, C. Marle, *Symplectic geometry and analytical mechanics*, Reidel, Dordrecht (1987).

[162] R. Lickorish, *An introduction to knot theory*, Graduate texts in mathematics **175**, New York, Springer (1997).

[163] T. Ligon, M. Schaaf, *On the Global Symmetry of the Classical Kepler Problem*, Reports on Math. Phys. **9** (1976), 281–300.

[164] G. Lion, M. Vergne, *The Weil representation, Maslov index and Theta Series*, Progress in Mathematics **6**, Birkhäuser (1980).

[165] J. Llibre, L. Roberto, *On the periodic orbits and the integrability of the regularized Hill lunar problem*, J. Math. Phys. **52**(8), 082701, 8 pp. (2011).

[166] G. Lodder, *Dihedral homology and the free loop space*, Proc. London Math. Soc. (3) **60** (1990), 201–224.

[167] R. Lohner, *Computation of guaranteed enclosures for the solutions of ordinary initial and boundary value problems*, Computational ordinary differential equations (London, 1989), 425–435, Inst. Math. Appl. Conf. Ser. New Ser., 39, Oxford Univ. Press, New York, 1992.

[168] Y. Long, *Index theory for symplectic paths with applications*, Progress in Mathematics, 207. Birkhäuser Verlag, Basel, 2002. xxiv+380 pp. ISBN: 3-7643-6647-8

[169] Y. Long, D. Zhang, C. Zhu, *Multiple brake orbits in bounded convex symmetric domains*, Adv. Math. **203** (2006), 568–635.

[170] J. Marsden, T. Ratiu, *Introduction to Mechanics and Symmetry*, 2nd edition, Text in Appl. Math. **17**, Springer Verlag (1999).

[171] T. Matukuma, *On the periodic orbits in Hill's case*, Proc. Imp. Acad. Japan **6** (1930), 6–8, **8** (1932), 147–150, **9** (1933), 364–366.

[172] D. McDuff, D. Salamon, *Introduction to Symplectic Topology* 2nd edition, Oxford University Press (1998).

[173] D. McDuff, D. Salamon, *J-holomorphic Curves and Symplectic Topology* 2nd edition, Amer. Math. Soc., Providence, RI (2012).

[174] R. McGehee, *Some homoclinic orbits for the restricted three-body problem*, Thesis (Ph.D.) – The University of Wisconsin – Madison. 1969. 63 pp.

[175] W. Merry, *On the Rabinowitz Floer homology of twisted cotangent bundles*, Calc. Var. Partial Differential Equations **42** (2011), no. 3-4, 355–404.

[176] K. Meyer, D. Schmidt, *Hill's lunar Equations and the Three-Body Problem*, J. Diff. Eq. **44**, (1982), 263–272.

[177] J. Milnor, *Topology from the Differential Viewpoint*, The University Press of Virginia (1965).

[178] J. Milnor, *On the geometry of the Kepler problem*, Amer. Math. Monthly **90**(6), (1983), 353–365.

[179] H. Mineur, *Sur les systèmes mécaniques admettant n intégrales premières uniformes et l'extension à ces systèmes de la méthode de quantification de Sommerfeld,* C. R. Acad. Sci., Paris **200** (1935), 1571–1573.

[180] H. Mineur, *Sur les systèmes mécanique dans lesquels figurent des paramètres fonctions du temps. Étude des systèmes admettant n intégrales premières*

uniformes en involution. Extension à ces systèmes des conditions de quantification de Bohr–Sommerfeld, Journal de l'École Polytechnique, Série III, 143ème année (1937), 173–191 and 237–270.

[181] R. Moeckel, *Lectures on central configurations*, downloaded from www-users.math.umn.edu/∼rmoeckel/notes/CentralConfigurations

[182] C. Moore, *Braids in Classical Gravity*, Physical Review Letters 70, 3675–3679 (1993).

[183] R. Moore, R. Kearfott, M. Cloud, *Introduction to interval analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. xii+223 pp. ISBN: 978-0-898716-69-6

[184] E. Mora, *Pseudoholomorphic cylinders in symplectisations*, Doctoral dissertation, New York University (2003).

[185] M. Morse, *Calculus of variations in the large*, American Mathematical Society Colloquium Publication **18**, Providence, RI (1934).

[186] J. Moser, *Regularization of Kepler's problem and the averaging method on a manifold*, Comm. Pure Appl. Math. **23** (1970), 609–636.

[187] J. Moser, *Periodic orbits near an equilibrium and a theorem by Alan Weinstein*, Comm. Pure Appl. Math. **29** (1976), 724–747.

[188] J. Moser, C. Siegel, *Lectures on celestial mechanics*, Die Grundlehren der mathematischen Wissenschaften **187**, Berlin, Springer (1971).

[189] J. Moser, E. Zehnder, *Notes on dynamical systems*, Courant lecture notes in mathematics **12**, New York, NY, Courant Inst. of Math. Sciences (2005).

[190] F. Moulton, *Periodic Orbits*, Carnegie Inst. of Washington, Washington, D.C. (1920).

[191] D. Mumford, J. Fogarty, F. Kirwan, *Geometric invariant theory*, vol. **34** of Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer, Berlin, 3. edition (1994).

[192] W. Neutsch, K. Scherer, *Celestial mechanics*, Mannheim, BI-Wiss.-Verl. (1992).

[193] P. Newstead, *Geometric invariant theory*, In *Moduli spaces and vector bundles*, vol. **359** of London Math. Soc. Lecture Note Ser., (2009), 99–127.

[194] D. Ó Mathúna, *Integrable systems in celestial mechanics*, Progress in mathematical physics **51**, Boston u. a., Birkhäuser (2008).

[195] J. Ortega, T. Ratiu, *Momentum maps and Hamiltonian reduction*, Progress in Math. **222**, Birkhäuser, Boston (2004).

[196] L. O'shea, R. Sjamaar, *Moment maps and Riemannian symmetric pairs*, Math. Ann. **31** (2000), 415–457.

[197] Y. Osipov, *Geometrical interpretation of Kepler's problem*, *Uspehi Mat. Nauk*, **27**(2) (1972), 161 (in Russian).

[198] Y. Osipov, *The Kepler problem and geodesic flows in spaces of constant curvature*, Celest. Mech. **16** (1977), 191–208.

[199] Y. Ostrover, *When symplectic topology meets Banach space geometry*, Proceedings of the ICM 2014, vol 2, 959–981.

[200] J. Palis, C. Pugh, *Fifty problems in dynamical systems*, in *Dynamical Systems*, Lect. Notes Math. **468** (1975), 345–353.

[201] W. Pauli, *Über das Modell des Wasserstoffmolekülions*, Annalen der Physik **68** (1922), 177–240.

[202] W. Pauli, *Über das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik*, Zeitschrift für Physik **36** (1926), 336–363.

[203] N. de Paulo, P. Salomão, *Systems of transversal sections near critical energy levels of Hamiltonian systems in* $\mathbb{R}^4$, arXiv:1310.8464

[204] H. Plummer, *On periodic orbits in the neighbourhood of centres of libration*, Monthly Notices Roy. Astron. Soc. **62**, 6 (1901).

[205] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* I–III, Gauthiers-Villars, Paris (1899).

[206] H. Poincaré, *Sur les lignes géodésiques des surfaces convexes*, Transaction Amer. Math. Soc. **6** (1905), 237–274.

[207] H. Poincaré, *Sur un théorème de géométrie*, Rend. Circ. Matem. Palermo **33** (1912), 375–407.

[208] C. Pugh, C. Robinson, *The* C^1 *closing lemma, including Hamiltonians*, Ergodic Theory Dynam. Systems **3** (1983), no. 2, 261–313.

[209] P. Rabinowitz, *Periodic solutions of Hamiltonian systems*, Comm. Pure Appl. Math. **31** (1978), 157–184.

[210] A. Robadey, *Différentes modalités de travail sur le général dans les recherches de Poincaré sur les systèmes dynamiques*, Thèse de Doctorat, Université Paris 7 Denis Diderot (2006).

[211] J. Robbin, D. Salamon, *The Maslov index for paths*, Topology **32** (1993), 827–844.

[212] J. Robbin, D. Salamon, *The spectral flow and the Maslov index*, Bull. L.M.S. **27** (1995), 1–33.

[213] C. Robinson, *A global approximation theorem for Hamiltonian systems*, Proc. Symp. Pure Math. **XIV**, Global Analysis, AMS (1970), 233–244.

[214] F. Schlenk, *Embedding problems in symplectic geometry,* de Gruyter Expositions in Mathematics **40**, Walter de Gruyter, Berlin (2005).

[215] F. Schlenk, *Symplectic embedding problems, old and new*, preprint.

[216] M. Schwarz, *Cohomology operations from* S^1*-Cobordism in Floer Homology*, Ph.D thesis, ETH Zürich (1996).

[217] M. Schwarz, *Equivalences for Morse Homology*, Geometry and Topology in Dynamics, Contemp. Math. **246**, AMS, Providence, RI (1999), 197–216.

[218] H. Seifert, *Periodische Bewegungen mechanischer Systeme*, Math. Z. **51** (1948), 197–216.

[219] R. Siefring, *Relative asymptotic behavior of pseudoholomorphic half-cylinders*, Comm. Pure Appl. Math. **61** (2008), no. 12, 1631–1684.

[220] R. Siefring, *Intersection theory of punctured pseudoholomorphic curves*, Geom. Topol. **15** (2011), no. 4, 2351–2457.

[221] R. Siefring, *Finite-energy pseudoholomorphic planes with multiple asymptotic limits* , arXiv:1607.00324

[222] C. Simó, T.J. Stuchi, *Central stable/unstable manifolds and the destruction of KAM tori in the planar Hill problem* Phys. D 140 (2000), no. 1-2, 1–32.

[223] C. Simó, *New families of solutions in N-body problems*, (English summary) European Congress of Mathematics, Vol. I (Barcelona, 2000), 101–115, Progr. Math., 201, Birkhäuser, Basel, 2001.

[224] J. Souriau, *Quantification géométrique. Application.*, Ann. Inst. H. Poincaré **6**, 311–341.

[225] J. Souriau, *Structures des Systèmes Dynamiques*, Dunod, Paris (1970).

[226] J. Souriau, *Sur la variété de Képler*, in *Convegno die Geometria Simplettica e Fisica Matematica,* Ist. Naz. di Alta Matematica, Roma 1973, Symposia Mathematica, Vol. XIV, Academic Press, London (1974).

[227] J. Souriau, *Géométrie globale du problème à deux corps*, Atti Accad. Sci. Torino CI. Sci. Fis. Mat. Natur. **117**(1) (1983), 369–418.

[228] S. Suhr, K. Zehmisch, *Linking and closed orbits*, Abh. Math. Semin. Univ. Hambg. 86 (2016), no. 1, 133–150.

[229] F. Spirig, J. Waldvogel, *Chaotic motion in Hill's lunar problem*, From Newton to chaos, 217–230, NATO Adv. Sci. Inst. Ser. B Phys. **336**, Plenum, New York (1995).

[230] E. Strömgren, *Connaissance actuelle des orbites dans le problème des trois corps*, Bull. Astron. **9** (1935), 87–130.

[231] V. Szebehely, *Theory of Orbits – The Restricted Problem of Three Bodies*, Academic Press, New York (1967).

[232] B. Szökefalvi-Nagy, *Spektraldarstellungen linearer Transformationen des Hilbertschen Raumes*, Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin: Springer (1942).

[233] C. Taubes, *The Seiberg–Witten equations and the Weinstein conjecture*, Geom. Topol. **11** (2007), 2117–2202.

[234] J. Thorpe, *Elementary Topics in Differential Geometry*, Undergraduate texts in mathematics, New York, Springer (1979).

[235] W. Tucker, *Validated numerics. A short introduction to rigorous computations.*Princeton University Press, Princeton, NJ, 2011. xii+138 pp. ISBN: 978-0-691-14781-9

[236] H. Varvoglis, C. Vozikis, K. Wodnar, *The two fixed centers: An exceptional integrable system*, Celestial Mech. Dyn. Astronomy **89** (2004), 343–356.

[237] C. Viterbo, *Metric and isoperimetric problems in symplectic geometry*, J. Amer. Math. Soc. **13**, no. 2, (2000), 411–431.

[238] T. Vozmischeva, *Integrable Problems of Celestial Mechanics in Spaces of Constant Curvature*, Kluwer Academic Publishers, Dordrecht (2003).

[239] A. Weinstein, *Symplectic manifolds and their lagrangian submanifolds*, Adv. in Math. **6** (1971), 329–346.

[240] A. Weinstein, *Lectures on Symplectic manifolds*, CBMS Conference Series **29**, American Mathematical Society, Providence, RI (1977).

[241] A. Weinstein, *On the hypothesis of Rabinowitz periodic orbit theorems*, J. Diff. Eq. **33** (1979), 353–358.

[242] A. Wintner, *Über die Konvergenzfragen der Mondtheorie*, Math. Z. **30** (1929), no. 1, 211–227.

[243] A. Wintner, *The Analytical Foundations of Celestial Mechanics*, Princeton University Press (1941).

[244] Z. Xia, *Melnikov Method and Transversal Homoclinic Points in the Restricted Three-Body Problem*, Jour. Diff. Eq. **96** (1992), 170–184.

[245] A. Zinger, *Pseudocycles and integral homology*, Trans. Amer. Math. Soc. **360** (2008), no. 5, 2741–2765.

Index

U. Frauenfelder, O. van Koert, *The Restricted Three-Body Problem and Holomorphic Curves*,
Pathways in Mathematics, https://doi.org/10.1007/978-3-319-72278-8