CHRISTOPHE GOLÉ

SYMPLECTIC TWIST MAPS

GLOBAL VARIATIONAL TECHNIQUES

FOREWORD

8/31/99

To the reader of this draft: this is a preliminary version of the book, which needs to undergo a good amount of revision. Comments such as this one appear at the top of each chapter to warn my early reader, and remind myself, of some of the work needed. My home-spun reference system is not completed yet. Chapters come with names as this point (because the numbers were variable). As an example, Proposition TOPOpropcz belongs to the appendix named TOPO and the number it corresponds to is at the end of that appendix. I have marked points that need my attention in the future by: ???. Where concepts are defined for the first time, I use the *definition style*. I will build an index listing all terms that appear in that font. This is not the format required by my publisher World Scientific. In the interest of saving paper, I kept the magnification low (the book would have 250 pages in the final format). As a result, some page transition could have been smoother. Finally, this version does not contains all the pictures.

Area preserving maps of the annulus first appeared in the work of Henri Poincaré (ref ???) on the three-body problem. As two dimensional discrete dynamical models, they offered a handle for the study of a complicated Hamiltonian system. Since then, these maps and their more specialized offspring called twist maps, have offered many opportunities for rigorous analysis of aspects of Hamiltonian systems, as well as an ideal test ground for important theories in that field (*eg.* Moser (1962) proved the first differentiable version of the KAM theorem in the context of twist maps).

This book is intended for graduate students and researchers in mathematics and mathematical physics interested in the interplay between the theories of twist maps and Hamiltonian dynamical systems. The original mandate of this book was to be an edited version of the author's thesis on periodic orbits of symplectic twist maps of $\mathbb{T}^n \times \mathbb{R}^n$. While it now comprises substantially more than that, the results presented, especially in the higher dimensional case, are still very much centered around the author's work.

At the turn of the 1980's, the theory of twist maps received a tremendous boost from the work of Aubry and Mather. Aubry, a solid-state physicist, had been led to twist maps in his work on ground states for the Frenkel-Kontorova model. This system, which models deposition on periodic 1–dimensional crystals, while not dynamical, provides a variational approach which is surprisingly relevant to twist maps. Mather, a mathematician who had worked on dynamical systems and singularity theory, gave a proof of existence of orbits of all rotation numbers in twist maps, what is now known as the Aubry-Mather theorem, using a different variational approach proposed by Percival. Aubry, who had conjectured the result, gave a proof using his approach. Both researchers then developed a sophisticated theory using an interplay of their two approaches. This lead to a flurry of work in mathematics and physics.

At about the same time, Conley & Zehnder (1983) gave a proof of the Arnold conjecture on the the torus, which heralded the birth of symplectic topology. This conjecture (now a theorem) states that the number of fixed point for a Hamiltonian map on a closed manifold is regimented in the same way as the number of critical points of real valued functions on that manifold. The proof involved Conley's generalized Morse theory for the study of the gradient flow of the Hamiltonian action functional in loop space. Later, with the influx of Gromov's holomorphic curve theory, this gave rise to Floer cohomology (Floer (???)). Interestingly, Arnold (1978) introduced his conjecture as a generalization of the famous fixed point theorem for annulus maps of Poincaré and Birkhoff, by gluing two annuli into a torus.

This book, while establishing a firm ground in the classical theory of twist maps, reaches out, via generalized symplectic twist maps, to Hamiltonian systems and symplectic topology. One of the approaches used throughout is that of the gradient flow of the action functional stemming from the twist maps' generating functions. We hope to convey that symplectic twist maps offers a relatively simple, often finite dimensional, interface to the variational and dynamical study of Hamiltonian systems on cotangent bundles.

Results for the two dimensional theory presented here include the classical theorems by Poincaré, Birkhoff (Chapter 7 and INVchapter), Aubry and Mather (Chapter AM). A joint work of the author with Sigurd Angenent on the vertical ordering of Aubry-Mather sets appears for the first time here (GCchapter). The approach of this book to the two dimensional theory is deliberately variational (except for Katznelson and Ornstein recent proof of Birkhoff's Graph Theorem in INVchapter) as I sought continuity between the low and high dimensions. Unfortunately, this choice leaves out the rich topological theory of twist maps and, more generally two dimensional topological dynamics. I refer the reader interested in the topological approach to Hall & Meyer (???), LeCalvez (1990) and the bibliography therein.

In higher dimensions, results by the author form the main focus of attention. These results are about existence of periodic orbits and their multiplicity for both symplectic twist maps and Hamiltonian systems on cotangent bundles (Chapter 4 and Chapter 7). The results on Hamiltonian systems use techniques of decompositions of these systems into symplectic twist maps . In Chapter 6, we provide the necessary connections between these maps and Hamiltonian and Lagrangian systems, some for the first time in the literature. In particular, M. Bialy and L. Polterovitch were kind (and patient!) enough to allow me to include their proof of suspension of a symplectic twist map by an optical Hamiltonian flow. Appendix 2 or TOPO establishes the parts of Conley's theory needed in the book, including some refinements that, to my knowledge, never appeared before. For readers uncomfortable with these topics, I try to motivate this appendix (chapter ???) by a hands–on introduction to homology and Morse theory. Chapter 9 presents Chaperon's proof of Arnold's conjecture on the torus, and the commonality between our methods and those of generating phases used in symplectic topology. Appendix 1 or SG , a self contained introduction of symplectic geometry, gathers (and proves most of) the results of symplectic geometry needed in the book.

The results in this book do not make minimizing orbits their central item. In fact, they often deliberately concern systems that cannot have minimizers (non positive definite twist). However, Chapter AMG is devoted to surveying the state of affairs in the generalizations of the Aubry-Mather theory to higher dimensions, where minimizers play a fundamental role. INVchapter, a poor substitute to a treatment that should occupy a volume on its own, surveys the theories of invariant tori (KAM theory and generalizations of Birkhoff's Graph Theorem by Bialy, Polterovitch and Herman), as well as that of splitting of separatrices.

The different topics in this book require different background from the reader. I have striven to make it possible for readers only interested in twist maps of the annulus or of $\mathbb{T}^n \times \mathbb{R}^n$ to read the sections pertaining to these topics with a minimum of reference to symplectic or Riemannian geometry, or to Conley's theory. On the other hand the appendices on symplectic geometry and topology are written, at least in part, with the novice in mind.

INTRODUCTION

$8/27/99$

In this introduction, we tell three mathematical stories which introduce themes that are interwoven throughout the book. The first one is the evolution of the dynamics of conservative systems (the standard map here) as one pertubs them away from completely integrable. The second story is about the relationship between Lagrangian or Hamiltonian systems and symplectic twist maps, illustrated here by the connection between the billiard map and the geodesic flow on a sphere. The third story relates Poincaré's last geometric theorem to symplectic topology.

1. Fall From Paradise

Consider the map $F_0 : \mathbb{R}^2 \mapsto \mathbb{R}^2$ given by:

$$F_0(x, y) = (x + y, y).$$

F_0 shears any vertical line $\{x = x_0\}$ into the line $\{y \mapsto (x_0 + y, y)\}$, of slope 1: as y increases, the image point moves to the right. We say that F_0 satisfies the *twist condition*. F_0 is linear with determinant 1 and hence is area preserving. Since $F_0(x+1, y) = F_0(x, y) + (1, 0)$, this map descends to a map f_0 of the cylinder $\mathbb{S}^1 \times \mathbb{R}$. There, the x variable is seen as an angle. f_0 is called an area preserving twist map of the cylinder, or twist map in short. See Chapter 1 for a more detailed definition of twist maps. The map f_0 has an additional property that makes it special among twist maps: it preserves each circle $\{y = y_c\}$, on which it induces a rotation of angle y_c (measured in fraction of circumference). We say that f_0 is *completely integrable*. Completely integrable maps are the paradise lost of mathematicians, physicists and astronomers. Not only are the dynamics of such maps entirely understood, but the invariance of each circle $\{y = y_c\}$ assures that no point drifts in the vertical direction. In their original celestial mechanics settings, twist maps appeared as local models of sections of the Hamiltonian flow around an elliptic periodic orbit. In this setting, this lack of drift means stability of the orbit (and by extension, one hoped to establish the stability of the solar system...). Nearby points stay nearby under iteration of the map. Of course "real" systems are rarely completely integrable. But one of the driving paradigms in the theory of Hamiltonian dynamics is the study of how one falls from this completely integrable paradise, and how many of its idyllic features survive the fall.

Falling is easy. Perturb F_0 ever so slightly into an F_ϵ:

$$F_\epsilon(x, y) = \left(x + y - \frac{\epsilon}{2\pi} sin(2\pi x),\ y - \frac{\epsilon}{2\pi} sin(2\pi x)\right),$$

called the *standard map*. As the reader may check, the vertical lines are still twisted to the right, and the area is still preserved under F_ϵ. Looking at the computer pictures of orbits of F_0 and F_ϵ in Figure 1. 1, we see what appear as invariant circles. We also see new features in the orbits of F_ϵ: some structures resembling collars of pearls (elliptic periodic orbits and their "islands"), interspersed with regions filled with clouds of points (chaos and diffusion due to intersecting stable and unstable manifolds of hyperbolic periodic orbits). We also see some "broken" circles made of dashed lines (Cantori or Aubry-Mather sets). These new features become more and more predominant as the value of ϵ increases: the elliptic islands bulge, the chaotic regions spread, and less and less circles appear unbroken. In fact, if $\epsilon \geq 4/3$, a theorem of Mather (1986) says that no invariant circle survives. However, the deep theory of Kolmogorov-Arnold-Moser (KAM, see INVchapter) implies that uncountably many invariant circles remain for small ϵ, those that have a very irrational rotation angle. In fact these circles occupy a set of large relative measure in the cylinder. A natural question arises: *what happens to invariant circles once they break?* The answer to this question, given by the Aubry-Mather theorem (see Chapter AM), is that invariant circles are replaced by invariant sets called Aubry-Mather sets whose orbits retain most of the features of those of invariant circles (cyclic order, Lipschitz graph regularity, rotation number and minimization of action). The Aubry-Mather sets with orbits of irrational rotation numbers form Cantor sets, sometimes called Cantori; those with rational rotation numbers usually contain hyperbolic periodic orbits and, depending on the authors' conventions, associated elliptic orbits. Of course the Aubry-Mather sets with their gaps form no topological obstruction to the vertical drift of orbits. In fact Mather (1991a) and Hall (1989) prove that, in a region with no invariant circle, one can find orbits visiting any prescribed sequence of Aubry-Mather sets. Hence these vestiges of stability have now become a stairway to drift and instability! The theory of transport (see Meiss (1992)) points at the regulatory role Aubry-Mather sets have on the *rate* of vertical diffusion of points.

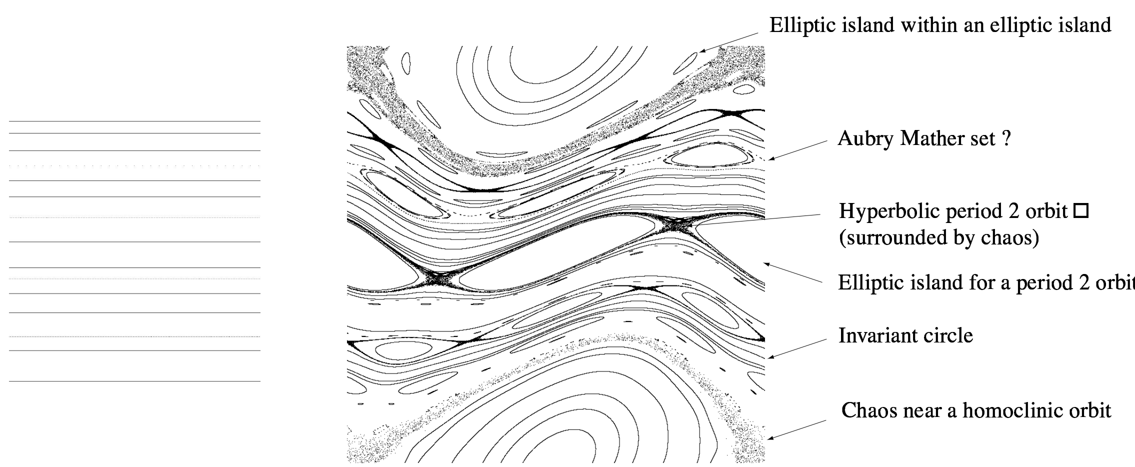


Fig. 1. 1. The different dynamics in the standard map: the left hand side shows a selection of orbits for the completely integrable F_0, whereas the right hand side displays orbits for F_ϵ with $\epsilon = .817$. Of course, one has to take computer generated figures with a grain of salt: computers cannot deal with irrational numbers...

Higher Dimensions

Make $F_0 : (x, y) \mapsto (x + y, y)$ defined above into a map of $\mathbb{R}^n \times \mathbb{R}^n$ by having x, y be vector variables. In analogy to the former situation, F_0 descends to a map f_0 from $\mathbb{T}^n \times \mathbb{R}^n$ to itself (x is now a vector of n angles). This space can be interpreted as the cotangent bundle of the torus, an important space in classical mechanics. Not only has the differential DF_0 determinant 1, but it also preserves the symplectic 2-form $\sum_k dx_k \wedge dy_k$ (the two notions were indistinguishable in dimension 2). The vertical fibers $\{x = x_c\}$ are still sheared, in a way made precise in Chapter STM. The map f_0 is called a *symplectic twist map* in this book. Our new f_0 is again called completely integrable as it preserves the tori $\{y = \mathrm{y_c}\}$, and induces a translation by the vector y_c on each one. One can perturb f_0 (in the realm of symplectic twist maps) and ask the same kind of questions as in the 2–dimensional case: what of the well understood, stable dynamics of f_0 survives a perturbation of the map, small or large?

It turns out that KAM theory still holds in this case, and guarantees the existence of many invariant tori whose dynamics is conjugated to the translation by (very) irrational vectors. One of the results central to this book is that for arbitrary perturbations, periodic orbits of any rational rotation vector exist for all symplectic twist maps of a large class, and a lower bound on their number is related to the topology of $\mathbb{T}^n$ (see Chapter 4). What about orbits of irrational rotation vector? Strictly speaking, there cannot be a $full$ analog of the Aubry-Mather theorem in higher dimensions. Mather (1991b) developed a powerful theory of minimal invariant measures and their rotation vectors on cotangent bundles of arbitrary compact manifolds. This theory proves the existence and regularity of many minimizing orbits. But in the case where the manifold is $\mathbb{T}^n$ with $n \geq 3$, the theory cannot guarantee that more than n directions be represented in the set of all rotation vectors of minimizing orbits. And indeed, some examples exist of maps (or Lagrangian systems) of $\mathbb{T}^3 \times \mathbb{R}^3$ all of whose recurrent minimizing orbits have rotation vector restricted to exactly 3 axes. If one lets go of the requirement that the orbits be action minimizers, then in certain examples, orbits of all rotation vectors can be found. The work of MacKay & Meiss (1992) points to a general theory for maps very far from integrable, but the case of maps moderately close to integrable, where less help from chaos can be expected, is not understood. Interestingly, if one trades the cotangent of a torus for that of a hyperbolic manifold, a large amount of the Aubry-Mather theory can be recovered: minimizing orbits of all rotation "direction", and of at least countably many possible speed in each direction exist (see Boyland & Golé (1996b)). Also, full fledge generalizations of the Aubry-Mather theorem exist in higher dimensional, but non dynamical settings generalizing the Frenkel-Kontorova model, as well as for some PDE's (de la Llave (1999)). We survey all these questions in greater detail in Chapter AMG.

2. Billiards and Broken Geodesics

Symplectic twist maps have rich ties with Hamiltonian and Lagrangian systems. They often appear as cross sections or discrete time snapshots of these systems. In Lagrangian systems, a trajectory γ is an extremal of an action functional $\int_\gamma Ldt$.In twist maps, this relates to an action function which is a discrete sum of the form $\sum S_k(x_k, x_{k+1})$ where x_k is a sequences of points of the configuration manifold and S_k are generating functions of twist maps. We explore this relationship in Chapter 6. A beautiful illustration of this occurs in the billiard map. The billiard we consider is planar, convex, and trajectories of a ball inside it are subject to the law of equality between angle of reflection and angle of incidence. Since we know that it is a straight line between rebounds, a trajectory is prescribed by one of its points of rebound and the angle of incidence at this rebound. In this way, we obtain a map $f : (x, y) \mapsto (X, Y)$, where x is the coordinate of the point of rebound and $y = -cos(\theta)$, where θ is the angle of incidence (see Figure 2. 1). Since x is the point of a (topological) circle, and y is in the interval $(-1, 1)$, the map f acts on the annulus $\mathbb{S}^1 \times (-1, 1)$. The choice of y instead of θ insures that f preserves the usual area in these coordinates (see Section TWISTsecexamples). The twist condition for f is a consequence of the convexity of the billiard: if one increases y ($i.e.$ increases θ) leaving x fixed, X increases.

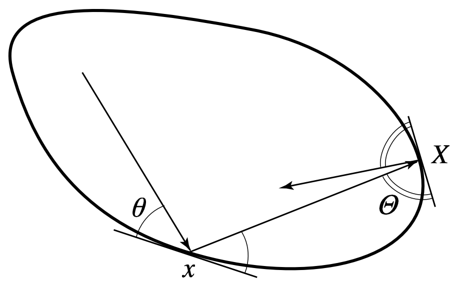


Fig. 2. 1. In a convex billiard, the point x and angle θ at a rebound uniquely and continuously determines the next point X and incidence angle Θ.

The map f can be seen as a limit of section maps for the geodesic flows of a sphere that is being flattened until front and back are indistinguishable. The boundary of the billiard is the (not so round in our illustration) fold of the flattened sphere. [To define the geodesic flow on the unit tangent bundle of the sphere, take a point on the sphere and a unit tangent vector (parameterized by its angle with respect to some tangent frame). Now travel at constant speed along the unique geodesic passing through this point and in the direction prescribed by the vector]. Draw on the sphere the closed curve C which eventually becomes the fold as one flattens the sphere. For a sufficiently flat sphere, all the geodesics on the sphere (except for maybe C, if it is a geodesic) eventually cross C transversally, and one can construct a section map which to one crossing at a certain point and angle makes correspond the next crossing point and angle. Seen in the three dimensional unit tangent bundle, the curve C lifts to a surface parameterized by points in C and all possible crossing angles in $(0, \pi)$, $i.e.$ an annulus, which all trajectories (except maybe for C) of the geodesic flow eventually cross transversally. [Poincaré initiated a similar section map construction in a 3–dimensional energy manifold for the restricted 3–body problem]. The annulus maps that one obtains in this fashion limit, as one flattens the sphere, to the billiard map. To see this, note that the geometry of the flat sphere near a point not on the fold is that of the Euclidean plane, where geodesics are straight lines. At a fold point, the law of reflexion is a simple consequence of what happens to a straight line segment as it is folded along a line transverse to it (see Figure 2. 2).

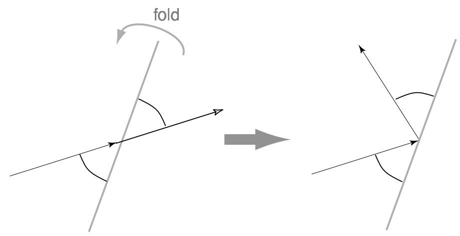


Fig. 2. 2. The law of reflexion as a consequence of folding.

Geodesics are length extremals among all (absolutely continuous) curves on the sphere. It therefore comes as no surprise that orbits of the billiard map are extremals of the length on the space of polygonal lines with vertices on the boundary (see Section TWISTsecexamples). If we inflate our billiard back a little, polygonal lines become *broken geodesics*. Indeed, the straight line segments can be replaced by segments of geodesic which, since the law of reflexion is not observed at a rebound for a general polygonal line, meet at an angle. In this space of broken geodesics, parameterized by the break points, geodesics are critical for the length function. To see why this is not only a beautiful, but also useful idea, consider the special case of periodic orbits of a certain period for the billiard map and geodesic flow. In the billiard, these correspond to closed polygons (see Figure 2. 3), parameterized by their vertices which form a *finite* dimensional space, whose topology clearly has to do with that of the circle. The same holds for geodesics of our almost flat sphere. In fact, when studying closed geodesics (or geodesic between two given points) on *any* compact manifold *one can restrict the analysis from the infinite dimensional loop space to a finite subspace of broken geodesics*. This was a key idea in Morse's analysis of the path space of a manifold (see Milnor (1969)). And, more generally applied to Hamiltonian systems, it is one of the important themes of this book: symplectic twist maps can be used to break down the infinite dimensional variational analysis of Hamiltonian systems to a finite dimensional one. This is discussed in detail in Chapter 6, and again in Chapter 9.

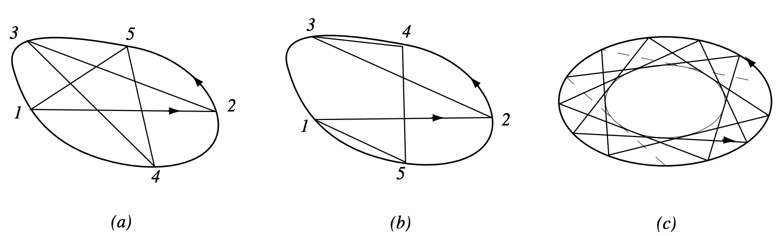


Fig. 2. 3. Different polygonal configurations in billiards: (a) is of period 5, rotation number 3/5 and is cyclically ordered. (b) is also of period 5, but of rotation 1/5 and is not cyclically ordered. Note that neither (a) nor (b) represent orbits since the law of reflexion is not satisfied. (c) is a configuration corresponding to an orbit on an invariant circle for the completely integrable elliptic billiard map. Its rotation number is presumably irrational.

Rotation Number and Ordered Configurations

The billiard map also provides a nice illustration of the notion of rotation number of periodic orbits (see Figure 2. 3 (a) and (b)). A consequence of the Aubry-Mather theorem is that any convex billiard has orbits of all rotation number in $(-1, 1)$. Polygonal curves corresponding to orbits on an invariant circle with irrational rotation numbers are all tangent to a circle or *caustic* inside the billiard (see Figure 2. 3 (c)). Polygonal curves corresponding to Aubry-Mather sets are "tangent" to a Cantor set. Finally, the billiard gives us an illustration of the notion of order for configurations of points. In Example (a) of Figure 2. 3, the configuration is *cyclically ordered*, in that the cyclic order of rebound points is conserved on the boundary after following them to their next rebound. Example (b) is, on the other hand not cyclically ordered. This notion of order is crucial to the Aubry-Mather theory. In Chapter AM, we present a proof of the Aubry-Mather theorem similar to that of Aubry's, in which one finds (cyclically ordered) orbits of irrational rotation numbers by taking limits of cyclically ordered periodic orbits. In GCchapter, we make use of the fact that the gradient flow for the action function (the length in the billiard map) of a twist map preserves the set of cyclically ordered configurations to give another proof of the Aubry-Mather theorem. We use a stronger order property of the flow as well in our proof that Aubry-Mather sets are vertically ordered. Unfortunately, there is no natural order for orbits of higher dimension twist maps. But the same kind of ordering exists in higher dimensional Frenkel-Kontorova models, for which the Aubry-Mather holds, as well as for certain PDEs (see de la Llave (1999)). The gradient flow in the PDE setting corresponds to generalized heat flows. The analogy to the preservation of order is given by theorems of comparison. This analogy, which was already noticed by Angenent (1988) , inspired him to introduce the gradient flow of the action in twist maps.

3. An Ancestor of Symplectic Topology

At the end of his life, Poincaré (1912) published a theorem, sometimes called his last geometric theorem, that can be simply stated as: *Let f be an area preserving map of a compact annulus, which moves points in opposite directions on the two boundary circles. Then f must have at least two fixed points.*

Poincaré gave an incomplete proof of this theorem, writing a moving letter of apology to the editor which mentions his bad health and expresses his desire that his work on this problem not be lost for posterity. Birkhoff gave a substantially different proof, which was also somewhat incomplete as to the existence of at least *two* fixed points (it did prove the existence of at least one). Since then, a number of proofs have appeared (Brown (???pb), Fathi (1983), Franks (1988), as well as Golé & Hall (1992), where the original proof of Poincaré is completed). We now sketch a proof of the theorem, in the very simple case where the map f also satisfies the twist condition. The ideas involved connect the original proof of Poincaré, the proof of LeCalvez (astérisque) we present in Section PBsecpb and the modern theory of symplectic topology.

Let F be the lift of f to the strip $\mathcal{A} = \{(x, y) \mid x \in \mathbb{R},\ y \in [0, 1]\}$, which moves boundary points in opposite directions. Such a lift always exists. Denote by (X, Y) the image of a point (x, y) by F. Consider

$$\Gamma = \{(x, y) \in \mathcal{A} \mid X(x, y) = x\},$$

which is the set of points that only move up or down under the map[1]. The twist condition means that the image of each vertical segment $\{x = x_0\}$ by F intersects that segment exactly at one point. This implies that

[1] Poincaré considered the similar set of points that only moved left or right, see Golé & Hall (1992)

Γ is a graph over the x–axis, and, by periodicity, the lift of a circle γ enclosing the annulus. Clearly, $f(\gamma)$ must also be a circle, graph over the x–circle. Any point in the intersection $\gamma \cap f(\gamma)$ is necessarily fixed by f: such points move neither left, right, nor up, nor down. This intersection is not empty, by area conservation. If $\gamma = f(\gamma)$ (as is the case if f is a completely integrable map), f has infinitely many fixed points. If not, area preservation dictates that there must be points of $f(\gamma)$ strictly above γ and others strictly below. Since both these sets are circles, this implies the existence of at least two points in the intersection, $i.e.$ two fixed points for f. □

We now show the connection between fixed points of f and critical points of a real valued function on the circle. As we will see in Chapter 1, the map F comes equipped with a generating function $S(x, X)$ which satisfies $S(x + 1, X + 1) = S(x, X)$ and $YdX - ydx = dS$. This derives directly from area preservation and conservation of boundaries. Consider the restriction w of S to Γ, $i.e.$ $w(x) = S(x, x)$. Write $\Gamma = \{(x, y(x))\}$ and $F(\Gamma) = \{(x, Y(x))\}$. By definition of Γ, $F(x, y(x)) = (x, Y(x))$. With this notation $dw = (Y(x) - y(x))dx$, which is zero exactly when $Y(x) = y(x)$:*the critical points of* w *correspond to intersections of* Γ *and its image by* F, *i.e. to fixed points of* F. By periodicity, w can be seen as a function of the circle, which must have a maximum and a minimum: two distinct critical points, unless w is constant, in which case all points of Γ must be fixed. This simple idea is key in Moser (1977), where it is shown that a generic symplectic maps has infinitely many periodic orbits around an elliptic fixed point. Arnold (1978) also motivates his famous conjecture on fixed points on symplectic manifolds by an argument similar to this one.

In the coordinates $(x, y') = (x, y - y(x))$, Γ becomes the 0–section $\{(x, 0)\}$, and $F(\Gamma) = \{(x, Y(x) - y(x))$ is the graph of the differential of w. The function w is called a generating (phase) function for the manifold $F(\Gamma)$.This is a simple instance of a more general situation: Γ and its image are Lagrangian manifolds, as is any 1–dimensional manifold in a 2–dimensional symplectic manifold (see Appendix 1 or SG). Important theorems in symplectic topology can be expressed, as this one, in terms of intersections of a Lagrangian manifold with the 0–section in some cotangent bundle. To find such intersections, one looks for critical points of generating phase functions for this manifold. As we have seen in the above example, it is easy to do so when the manifold is a graph over the 0–section. The first challenge is to deal with cases where a Lagrangian manifold is not a graph [This will occur for our sets Γ and $F(\Gamma)$ when the map f is not twist, for example]. One then seeks generating phase functions with extra variables [This is in effect what the proof of LeCalvez does: one obtains a generating phase function for the set $F(\Gamma)$ by adding the generating functions of the twist maps that decompose f]. The second challenge is to show that these general generating phase functions have the requisite number of critical points. This is done in this book using Conley's theory on the gradient flow of the generating phase function. One difficulty arises from the non compactness of the space on which this function is defined. One resolves that by seeking compact invariant set of a sufficiently complicated topology for the gradient flow. We called some of these sets "ghost tori" in Golé (1989). These sets have their analogs in the sets of connecting orbits between critical points of the action functional in loop space that Floer based his cohomology on. Although implicit in several parts of the book, we will not use the language of Lagrangian intersection and generating phase function before Chapter 9.

Section INTROsecpb is 3.0

CHAPTER 1 or TWIST

TWIST MAPS OF THE ANNULUS

$1/25/2000$

Action to be taken: Make and add figures. Correct typos. Revise after writing the intro (some of its material might be used in the intro: standard map, billiard, some definition of symplectic...) The section on elliptic fp may be spread over to Chapter SG. I just moved the section on Poincaré-Birkhofffrom the now defunct Chapter PB. Make sure the transition is smooth. Check out the background, and give reference about Poincaré sections and twist maps.

4. Monotone Twist Maps of the Annulus

A. Definitions

In the first part of this book, we consider diffeomorphisms of the annulus, or of the cylinder. The *annulus* can be defined as

$$\mathbf{A} = \mathbb{S}^1 \times [a, b].$$

[More generally, we could define $\mathbf{A} := \{(x, y) \in \mathbb{S}^1 \times \mathbb{R} \mid u_-(x) \le y \le u_+(x)\}$, where both u_- and u_+ smooth functions on $\mathbb{S}^1$]. We define the *cylinder* by:

$$\mathcal{C} = \mathbb{S}^1 \times \mathbb{R}.$$

As with maps of the circle, it is often less ambiguous to work with lifts of diffeomorphisms of $\mathbf{A}$. These are maps of the *strip*:

$$\mathcal{A} := \{(x, y) \in \mathbb{R}^2 \mid a \le y \le b\}$$

where x, thought of as the angular variable, ranges over $\mathbb{R}$. The *covering map* $proj : \mathcal{A} \mapsto \mathbf{A}$ takes (x, y) to $(x \mod 1\ , y)$ and a *lift* of a map f of the annulus is a map F of the strip which satisfies:

$$proj \circ F = f \circ proj.$$

This implies in particular that $F(x + 1, y) = F(x, y) + (n, 0)$, for some integer n. By continuity, n does not depend on the point (x, y), nor on the lift F of f, it is called the *degree* of f. In this book, we assume that f is an orientation preserving diffeomorphism of the annulus. In this case, the degree of f is 1 and

$$F(x + 1, y) = F(x, y) + (1, 0) \tag{4.1}$$

for any lift F of f. Denoting by T the translation $T(x, y) = (x + 1, y)$, equality (4.1) reads:

$$F \circ T = T \circ F \tag{4.2}$$

Clearly, any map F of $\mathcal{A}$ that satisfies (4.2) is the lift of a map f of $\mathbf{A}$ which has degree 1. We say that f is *induced* by F.

Definition 4.1 Let F be a diffeomorphism of $\mathcal{A} = \mathbb{R} \times [a, b]$ and write $(X(x, y), Y(x, y)) = F(x, y)$. Let F satisfy:
(1) *F preserves the boundaries of $\mathcal{A}$*: $Y(x, a) = a, Y(x, b) = b$.
(2) *Twist Condition:* the function $y \mapsto X(x_0, y)$ is strictly monotone for each given x_0.
(3) *Area and Orientation Preserving:* $\det DF = 1$ or, equivalently, $dY \wedge dX = dy \wedge dx$.
(4) $F \circ T = T \circ F$

Then F induces a map f on the annulus $\mathbf{A}$ which is called a *(area preserving, monotone) twist map of the annulus.*

Exercise 4.1 Prove the above statements about the degree of a map and its lifts.

B. Comments on the Definition

Twist Condition. Condition (2) implies that the map $y \mapsto X(x_0, y)$ is a diffeomorphism between the vertical *fiber* $\{x = x_0\}$ and its image on the x-axis (also called the *base*). In other words, the image of the fiber x_0 by F forms a graph over the x-axis, as is shown in Figure 4. 1.

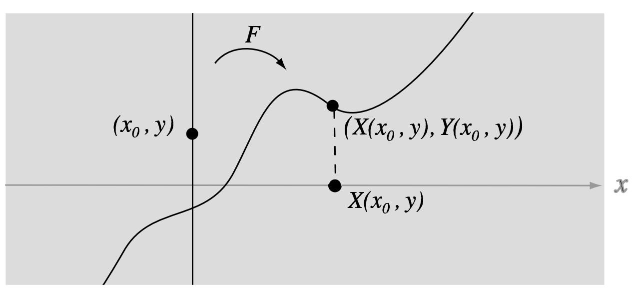


Fig. 4. 1. The positive twist condition: as one moves up along a vertical fiber, the image point moves right.

Often, the monotonicity of the map $y \mapsto X(x_0, y)$ is expressed by the equivalent derivative condition:

$$\frac{\partial X}{\partial y} \neq 0. \tag{4.3}$$

Since $\mathcal{A}$ is connected, this derivative is either always strictly positive, or always strictly negative. We say that F is a *positive twist map* (resp. *negative twist map*) if $y \mapsto X(x_0, y)$ is strictly increasing (resp. decreasing). Note that the lift of a positive twist map "moves" points on the upper boundary of $\mathcal{A}$ "faster" than on the lower boundary. If F satisfies the latter, we say that it has the *boundary twist condition.* This condition, much

weaker than the twist condition in Defi d in the Poincaré-Birkhoff theorem, see
TWISTsecpb.

We now show that the twist conditi $, y) \mapsto (x, X)$ is an *embedding* of $\mathcal{A}$ in
$\mathbb{R}^2$, *i.e.* a local diffeomorphism which rential of ψ is given by :

whose determinant $\frac{\partial X}{\partial y}$ is non zero by s a local diffeomorphism. To show that
it is injective, suppose $\psi(x_1, y_1) = \psi($ $_2$, and y_1 and y_2 are forced to be equal
because the map $y \mapsto X(x_1, y)$ is strict he reader to verify that, conversely, if ψ
is an embedding of $\mathcal{A}$, then the twist co as a change of coordinates.

Area Preservation, Flux and S of variable formula in multivariate inte-
gration shows that the infinitesimal cond $ea(X) = Area(F(X))$ for any domain
X in $\mathcal{A}$ (or for any Lebesgue measurab a preservation to another global notion:
that of flux. For F an area preserving m define the function $S : \mathcal{A} \to \mathbb{R}$ by:

where this path integral is over *any* cu nt z_0 and the variable $z = (x, y)$. Using
Stokes' theorem and Condition (3) and nected, one shows that S is well defined
on $\mathcal{A}$ (*i.e.* it is independent of the path $- ydx = dS$ (see Exercise 4.2). [This
is just an expression of the fact that clo mply connected regions]. The *flux* of an
area preserving map F of $\mathbb{R}^2$ satisfying :

This makes sense, since, by Stokes' t (see Exercise 4.3). The flux of the map
F can be seen geometrically in the cylin d between an embedded circle wrapping
once around $\mathcal{C}$ and its image by the m e 4. 2). Indeed, $S(\beta(1)) - S(\beta(0)) =$
$\int_\beta YdX - ydx = \int_{F(\beta)} ydx - \int_\beta ydx$ Now take β such that $\beta(1) = T\beta(0)$.

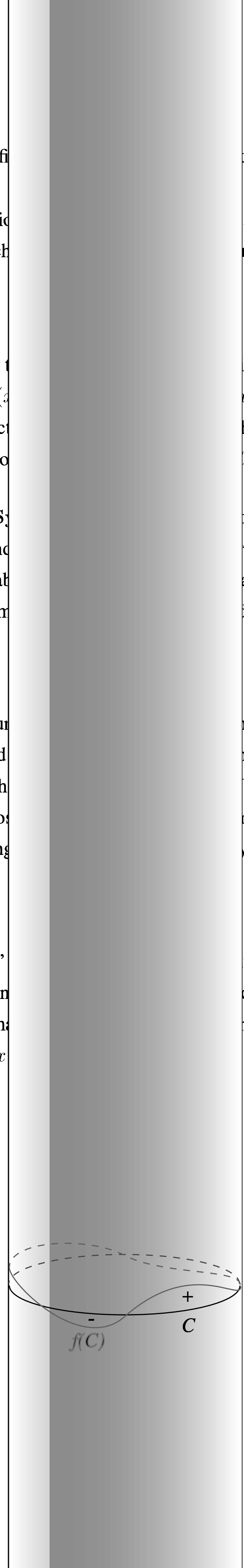


Fig. 4. 2. The flux of a cylinder map as the net area between an enclosing circle C and its image $f(C)$

If F preserves the boundary of a bounded strip $\mathcal{A}$, then f preserves the boundary circles and the flux is by force zero. When no such curve is preserved, the flux can take any value in $\mathbb{R}$ as the example $V_a(x,y) = (x, y+a)$ with $Flux(F) = a$ shows. Since examples of this type show no recurrent dynamics, we exclude them from our study by always imposing, directly or indirectly, the zero flux condition on our maps. If F has zero flux, then $S \circ T = S$ and thus S induces a function s on $\mathbf{A}$ such that

$$f^* ydx - yd = ds. \tag{4.4}$$

taking the exterior derivative on both sides of this equation, one gets $d(f^* ydx - ydx) = d^2 s = 0$, and thus

$$f^*(dy \wedge dx) = dy \wedge dx.$$

A map that satisfies this last equality is called *symplectic*, because it preserves the *symplectic form* is called *exact symplectic*. Hence (4.4) shows that exact symplectic implies symplectic. Hence if F has zero flux, the map f it induces is exact symplectic. Conversely, by Stokes' theorem, if f is exact symplectic, any of its lifts has zero flux (Exercise 4.2). Hence the map V_a of the cylinder defined above is not exact symplectic, even though it is symplectic. Note that, in contrast, a symplectic map F of the plane is always exact symplectic: as any closed form on the plane, $F^*(y \wedge dx) - ydx$ is exact (Poincaré's Lemma).

Exercise 4.2 a) Using Stokes Theorem, show that if λ is a closed 1–form on a simply connected domain of $\mathbb{R}^2$, then the function $S = \int_{z_0}^{z} \lambda$ is well defined (*i.e.* does not depend on the path of integration between z and z_0) and that $dS = \lambda$. Apply this to $\lambda = YdX - ydx$.
b) What should a definition of S be if F preserves a smooth area form $\alpha(x,y)dy \wedge dx$?

Exercise 4.3 a) Let F be an area preserving map of $\mathbb{R}^2$ with $F \circ T = T \circ F$. Show that for the function S defined above, $S \circ T - S$ is constant, and hence $Flux(F)$ is well defined. (*Hint.* Given two points z_1, z_2 in $\mathcal{A}$, take any two curves γ_1, γ_2, with γ_i joining z_i and Tz_i, $i = 1, 2$. Take a curve β joining z_1 and z_2 and apply Stokes Theorem to the closed curve $\beta \cdot \gamma_1 \cdot (T\beta)^{-1} \cdot \gamma_2^{-1}$.)
b) Show that any lift of an exact symplectic map of the cylinder has zero flux.
c) (For those who know about DeRham cohomology) Prove that $Flux(F)$ is the result of the pairing of the class $[f^* ydx - ydx]$ in $H^1_{DR}(\mathcal{C})$ with the first homology class represented by a circle going around the cylinder once in the positive direction (as usual, f is the map induced by F).

C. Twist Maps of the Cylinder

The comments of the previous subsection motivate the following:

Definition 4.4 (Twist Maps of the Cylinder) Let F be a diffeomorphism of $\mathbb{R}^2$ and write $(X(x,y), Y(x,y)) = F(x,y)$. Let F satisfy:

(1) *F is isotopic to the Identity*
(2) *Twist Condition:* the map $\psi := (x,y) \mapsto (x, X(x,y))$ is a diffeomorphism of $\mathbb{R}^2$
(3) *Area Preserving & Zero Flux (Exact Symplectic):* $YdX - ydx = dS$ with some real valued function S on $\mathbb{R}^2$ satisfying:

$$S(x+1,y) = S(x,y).$$

Then F is the lift of a map f on the cylinder $\mathcal{C}$ which is called a *monotone twist map of the cylinder.*

Condition (1) means that F can be deformed continuously into the identity through a path of homeomorphisms of the cylinder. For maps of the closed strip $\mathrm{I\!R} \times [a,b]$, this condition clearly implies that the boundaries have to be preserved, and hence Condition (1) here is the analog to Condition (19) in Definition 4.1. It will appear clearly in next section that the periodicity of the function S implies the periodicity $F \circ T = T \circ F$, $i.e.$ Condition (4) of Definition 4.1, which is necessary for F to induce a map of the cylinder. Finally, the condition that ψ be a diffeomorphism here can be relaxed: one can require that ψ only be an embedding, $i.e.$ a diffeomorphism of $\mathrm{I\!R}^2$ into a proper subset of $\mathrm{I\!R}^2$, to the cost of some (manageable) complications.

Remark 4.5 There exist several other definitions of monotone twist maps in the literature. Most noteworthy are the topological definitions, where the map is only required to be a homeomorphism (and not necessarily a diffeomorphism). The twist condition takes different forms with different authors. One commonly used is that the map $y \mapsto X(x,y)$ be monotonic (Boyland (1988), Hall (1984), Katok (1982), LeCalvez (astérisque)). A much milder condition is considered in Frank (1988), where certain neighborhoods must move in opposite directions around the annulus. The preservation of area is sometimes discarded by these authors, replaced by a condition that the map contracts the area, or that it is topologically recurrent. The topological theory for twist maps is extremely rich and would be the subject of an entire book. Our choice of working in the differentiable category stems from the possibilities of generalization to higher dimensions that it offers.

Exercise 4.3 Show that a map of the bounded annulus which is homotopic to Id preserves each boundary component (Note: the converse is also true, but much harder to prove).

5. Generating Functions and the Variational Setting

A. Generating Functions

In the previous section, we have seen that the lift F of a twist map of either the cylinder or the annulus comes with a function S such that $F^*ydx - ydx = YdX - ydx = dS$ and $S(x+1,y) = S(x,y)$. The first equation expresses the fact that F preserves the area, whereas the periodicity of S, expresses the zero flux condition.

On the other hand, the twist condition on F gives us a function ψ which we view as a change of coordinates $\psi : (x,y) \mapsto (x,X)$. In the (x,X) coordinates[2] the equation $YdX - ydx = dS(x,X)$ implies immediately that the functions $-y(x,X)$ and $Y(x,X)$ are the partial derivatives of S:

$$y = -\frac{\partial S(x,X)}{\partial x}, \qquad Y = \frac{\partial S(x,X)}{\partial X} \tag{5.1}$$

These simple equations are the cornerstone of this book. The function $S(x,X)$ is called the *generating function* of F in that from S we can retrieve F, at least implicitly: ψ^{-1} is given by $(x,X) \mapsto (x, -\frac{\partial S}{\partial x})$ hence ψ is implicitly given by S. Thus F is defined by:

[2] Remember that under the change of coordinates ψ, a function S changes according to $S \mapsto S \circ \psi$. Likewise, $y \mapsto y \circ \psi$ and $Y \mapsto Y \circ \psi$.

$$F : (x, y) \mapsto (X \circ \psi(x, y), \frac{\partial S}{\partial X}(\psi(x, y))) \tag{5.2}$$

and the two coordinates of F are given implicitly by the function S and its partial derivatives. In Proposition PROPgfstm of Chapter STM, we give conditions under which a function on $\mathbb{R}^2$ is a generating function of the lift F of some twist map. We also show that the correspondence between maps and their generating functions (mod constant) is one to one and continuous. The following exercise gives two necessary conditions for a function to generate a twist map:

Exercise 5.1 Show that if $S(x, X)$ is the generating function of a positive twist map, then:
a) $\partial_{12}S(x, X) < 0$
b) $S(x + 1, X + 1) = S(x, X)$

Exercise 5.2 Show that if F the lift of a twist map of the annulus $\mathbb{S}^1 \times [0, 1]$ then $S(x, X)$ can be interpreted as the area of the triangular shaped area with vertices $(x, 0)$, $(X, 0)$ and (X, Y) shown in Figure 5. 1. (*Hint.* Show geometrically on this picture that $Y = \frac{\partial S}{\partial X}$. For $y = -\frac{\partial S}{\partial x}$, consider the preimage of this triangular region by F). Solve question b) of the previous exercise using this geometric construction.

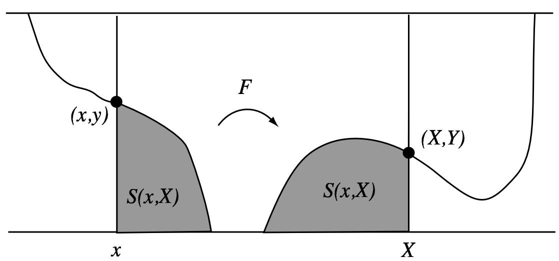


Fig. 5. 1. The generating function as an area

Exercise 5.3 Show that the inverse of a positive twist map with generating function $S(x, X)$ is a negative twist map with generating function $-S(X, x)$.

B. Variational Principle

The lift F of a twist map gives rise to a dynamical system whose orbits are given by the images of points of $\mathbb{R}^2$ under the successive iterates of F. The *orbit* of the point (x_0, y_0) is the biinfinite sequence:

$$\{\dots (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1), \dots, (x_k, y_k) \dots\}$$

where $(x_k, y_k) = f(x_{k-1}, y_{k-1})$.

Lemma 5.1 *Let F be a monotone twist map of $\mathcal{A}$ or $\mathbb{R}^2$ and let $S(x, X)$ be its generating function. There is a one to one correspondence between orbits $\{(x_k, y_k) = f^k(x_0, y_0)\}_{k\in\mathbb{Z}}$ of F and sequences $\{x_k\}_{k\in\mathbb{Z}}$ satisfying:*

$$\partial_1 S(x_k, x_{k+1}) + \partial_2 S(x_{k-1}, x_k) = 0 \qquad \forall k \in \mathbb{Z}. \tag{5.3}$$

The correspondence is given by: $y_k = -\partial_1 S(x_k, x_{k+1})$.

Proof. Let $\{(x_k, y_k)\}_{k\in\mathbb{Z}}$ be an orbit of F. Since $(x_k, y_k) = f(x_{k-1}, y_{k-1})$ for all integer k, Equation (5.1) implies:

$$y_k = -\partial_1 S(x_k, x_{k+1}) = \partial_2 S(x_{k-1}, x_k).$$

Conversely, let $\{x_k\}_{k\in\mathbb{Z}}$ satisfy Equation (5.3) and set $y_k = -\partial_1 S(x_k, x_{k+1})$, for all integer k. Then, applying Equations (5.2) and (5.3) :

$$\begin{aligned} f(x_{k-1}, y_{k-1}) &= f \circ \psi^{-1}(x_{k-1}, x_k) = (x_k, \partial_2 S(x_{k-1}, x_k)) \\ &= (x_k, -\partial_1 S(x_k, x_{k+1})) = (x_k, y_k). \end{aligned}$$

□

Equations (5.3) can be interpreted as "discrete Euler-Lagrange" equations for some action function on the space of sequences. Indeed, let F be the lift of a twist map of the cylinder, and $S(x, X)$ its generating function. Given a sequence of points $\{x_N, \ldots, x_M\}$, we can associate its *action* defined by:

$$W(x_N, \ldots, x_M) = \sum_{k=N}^{M-1} S(x_k, x_{k+1})$$

Corollary 5.2 (Critical Action Principle) *A sequence $\{x_N, \ldots, x_M\}$ is the projection of an orbit segment of F on the x-axis if and only if it is a critical point of W restricted to the subspace of sequences $\{w_N, \ldots, w_M\}$ with fixed endpoints: $w_N = x_N, w_M = x_M$.*

Proof. Given a sequence $\{x_N, \ldots, x_M\}$, introduce the sequences

$$y_k = -\partial_1 S(x_k, x_{k+1}) \quad \text{and} \quad Y_k = \partial_2 S(x_k, x_{k-1}).$$

In particular, $F(x_k, y_k) = (x_{k+1}, Y_k)$. If $\tilde{W}$ is the restriction of W to the set of sequences with fixed endpoints x_N and x_M, a direct calculation yields:

$$d\tilde{W}(x_N, \ldots, x_M) = \sum_{k=N+1}^{M-1} (Y_{k-1} - y_k) dx_k.$$

Hence $\{x_N, \ldots, x_M\}$ is a critical point for W if and only if $Y_{k-1} = y_k$, which is a rephrasing of Equation (5.3) , *i.e.* the sequence $\{(x_N, y_N), \ldots, (x_M, y_M)\}$ is an orbit segment. □

Exercise 5.4 Adapt Lemma 5.1 to a situation where the map F is a composition of different twist maps $F = F_k \circ \ldots \circ F_1$ with generating functions $S_1, \ldots, S_k$. Note that you do not need to assume that all the F_i are either positive twist (or all negative twist). If they are, one calls F a positive (*resp.* negative) *tilt map*.

C. Periodic Orbits

Let F be the lift of a twist map f of the annulus $\mathbf{A}$, or cylinder $\mathcal{C}$. Suppose that some orbit $\{x_k, y_k\}_{k\in\mathbb{Z}}$ of F satisfies:

$$x_{k+n} = x_k + m \tag{5.4}$$

that is, $F^n(x_k, y_k) = T^m(x_k, y_k)$. Then $f^n(proj(x_k, y_k)) = proj(x_k, y_k)$, and thus the orbit of (x_0, y_0) is the lift of a periodic orbit of f. We say that a sequence $\{x_k\}$ satisfying (5.4) is a (m, n) *sequence*. An orbit whose x projection is an (m, n) sequence is called a (m, n) *orbit*, or an orbit of *type* (m, n). Hence, under n iterates of F, points in a (m, n) orbit get translated by the integer m in the x direction. Down in the annulus, this can be interpreted as the orbit wrapping m times around the annulus in n iterates. Conversely, it is not hard to see that any periodic orbit of f of period n lifts to an (m,n) orbit of a lift F, for some integer m which *does* depend on the choice of F. The proof of the following is identical to that of Corollary 5.2:

Proposition 5.3 *A (m, n) periodic sequence is the x- projection of a m, n periodic orbit if and only if its is a critical point of $W(x_k, \dots, x_{k+q}) = \sum_{j=k}^{k+q-1} S(x_j, x_{j+1})$ for one (and hence for all) $k \in \mathbb{Z}$.*

Exercise 5.5 Show by an example that the number m for a periodic orbit of a twist map depends on the lift.

D. Rotation Numbers

Another interpretation of the numbers m, n in a periodic orbit is that the average displacement in the x direction of the points in a (m, n) orbit is m/n. In general, if $\{x_k, y_k\}_{k\in\mathbb{Z}}$ is *any* orbit, one can try to compute the limits:

$$\lim_{k\to+\infty} \frac{x_k}{k}, \qquad \lim_{k\to-\infty} \frac{x_k}{k}$$

If these limits exist, they are called respectively the *forward* and *backward rotation numbers*. If they are equal, they are called the *rotation number*. Since $\lim_{k\to\infty} \frac{x_k}{k} = lim_{k\to\infty} \frac{x_k - x_0}{k}$, the rotation number is an asymptotic measure of the average displacement in the x direction along an orbit. Obviously, an (m, n) periodic orbit has rotation number m/n. We also call *rotation number of the point* $z = (x, y)$ the rotation number of its orbit under F; we denote this number by $\rho_f(z)$.

Exercise 5.6 For those who know Birkhoff's ergodic theorem, show that, if f is an area preserving map of the annulus, $\rho_f(z)$ exists for a set of points z of full Lebesgue measure in $\mathcal{A}$ (*Hint.* $\lim \frac{x_k - x_0}{k} = \lim \frac{1}{k}\sum_1^k (x_j - x_{j-1})$ is the time average of some function).

6. Examples

A. The Standard Map

As noted in the introduction, one of the most widely studied family of monotone twist maps is the so called standard family, or *standard map*. We show how to retrieve explicitly the standard map from its generating function. Let

$$S(x,X)=\frac{1}{2}(X-x)^2+V(x),$$

where V is 1-periodic in x. Define

$$\begin{aligned} y &= -\partial_1 S(x,X) = X-x+V'(x) \\ Y &= \partial_2 S(x,X) = X-x. \end{aligned}$$

then it is easily seen that

$$\begin{aligned} X &= x+Y \\ Y &= y+V'(x), \end{aligned}$$

That is, S generates the lift of a twist map:

$$F(x,y)=(X,Y)=(x+y+V'(x),y+V'(x)).$$

Taking as "potential" V the 1-parameter family $\frac{k}{4\pi^2}cos(2\pi x)$, we do indeed get the standard family:

$$F_k(x,y)=(x+y-\frac{k}{2\pi}sin(2\pi x),y-\frac{k}{2\pi}sin(2\pi x))$$

When $V\equiv 0$ (or k is equal to 0 in the standard family), the generating function is $\frac{1}{2}(X-x)^2=\frac{1}{2}\mathrm{Dis}^2(x,X)$ and the map it generates is the *shear map*:

$$F_0(x,y)=(x+y,y)$$

which is *completely integrable*, in the sense that each horizontal line $\{y=y_0\}$ (covering a circle in $\mathcal{C}$) is invariant under F_0, and that the restriction of F_0 to $\{y=y_0\}$ is a translation: $x\mapsto x+y_0$ (lift of a rotation of angle $2\pi y_0$). We will see in Chapter HAM that F_0 is the time 1 map of the geodesic flow for the Euclidean metric on the circle.

As noted in the introduction, an important question about the standard family (or any set of maps containing a completely integrable one) is: which features of F_0 survive as one perturbs the parameter k away from 0?

Exercise 6.1 Check all the axioms of twist maps of the cylinder on the standard map.

B. Elliptic Fixed Points of Area Preserving Maps

The study of the dynamics around conservative elliptic fixed points was the motivation behind the birth of twist maps. It started when Poincaré studied the dynamics around an elliptic periodic orbit in the restricted 3–body problem. This is a Hamiltonian system (see Chapter SG) with 2 degrees of freedom, whose energy surface is 3–dimensional. Poincaré considered the return map on a 2–dimensional transverse section to the periodic orbit. Since the system is Hamiltonian, the return map is symplectic (see Theorem THMhamsym of Chapter SG). Generically, it is also shown to satisfy a twist condition. To formalize this a little, we present here the Birkhoff Normal Form Theorem. Poincaré was interested in proving that an elliptic periodic orbit is stable (leading to the more difficult question of the stability of the solar system), and in finding many periodic orbits close by. Both these problems were solved affirmatively for generic maps, the first by the KAM theory (see INVchapter) and the second by the theorem of Poincaré-Birkhoff (see TWISTsecpb).

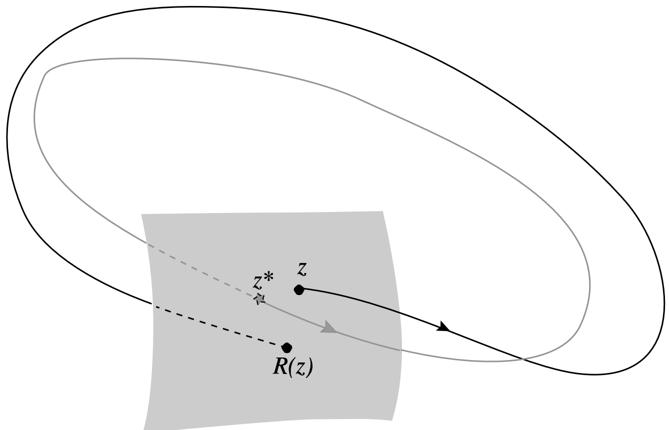


Fig. 6. 2. A Poincaré section around the periodic orbit of the point z^*, with the return map R.

Let F be a symplectic C^∞ diffeomorphism in a neighborhood of 0 in $\mathbb{R}^2$, which has 0 as a fixed point. Since $\det\ Df(0) = 1$, the two eigenvalues are either real $\lambda, 1/\lambda$ or complex $\lambda, \overline{\lambda}$ and conjugated on the unit circle. In the first case, we say that 0 is a *hyperbolic fixed point*, in the second case that it is an *elliptic fixed point* (see also Appendix 1 or SG). If F is the return map of a periodic orbit based at z^* as above, the periodic orbit is called elliptic or (resp. hyperbolic) when z^* is an elliptic (resp. hyperbolic) fixed point for R.

Suppose now that 0 is an elliptic fixed point and that $Df(0)$ has eigenvalues $\lambda = e^{i2\pi\alpha}$ and $\overline{\lambda}$ (i.e. Df(0) is a rotation of angle α). Suppose moreover that $\lambda^n \neq 1$ for n in $\{1, \ldots, q\}$ for some integer q. We can make a change of variable $z = x + iy, \overline{z} = x - iy$ and write the Taylor expansion of order n of $F(z)$ in these coordinates:

$$f(z) = \sum_{k=1}^{n} R_k(z, \overline{z}) + o(|z|^n)$$

Theorem 6.1 (Birkhoff Normal Form) *There exists a symplectic (for the form $dx \wedge dy$), C^∞ diffeomorphism h, defined near 0 and having 0 as a fixed point such that:*

$$h \circ f \circ h^{-1}(z) = \lambda z e^{i2\pi P(z\overline{z})} + o(|z|^{q-1})$$

or, in polar coordinates ($z = re^{i2\pi\theta}$):

$$\tilde{f} = \overline{h} \circ f \circ \overline{h}^{-1}(r, \theta) = \big(\theta + \alpha + P(r^2) + o(|r|^{2n}),\ r + o(|r|^{2n})\big)$$

where $P(x) = a_1 x + \ldots + a_m x^m$ with $2m + 1 < q$. Each of the "Birkhoff invariants" a_k is generically non zero.

For a proof of this, we refer to LeCalvez (1990) . There are also versions that require less differentiability (see Moser (1973)). The point of this theorem is that, if we make the generic assumption that some a_k is non zero, the map F satisfies a twist condition in a neighborhood of $r = 0$ (for $r > 0$). Note that, in polar coordinates, the map $\tilde{f}$ preserves the form $rd\theta \wedge dr$, (which is only non–degenerate for $r > 0$. By making a further change of variables that preserves the vertical foliation $\{x = ct\}$, one can get a map that preserves $d\theta \wedge dr$ (see Chenciner (1985)). This last map preserves no boundaries. However, one can extend it to a boundary preserving map of a compact annulus. The main results in the theory can often be made precise enough to tell apart the dynamics of the original map from that of the extension. Hence the dynamical study around conservative fixed points reduces to the study of twist maps.

C. The Frenkel-Kontorova Model

The variational approach in Section 5 was encountered by Aubry (see Aubry & Le Daeron (1983)) while studying a model in condensed matter physics. In this model, one considers a chain of particles whose nearest neighbor interaction is represented by springs. The chain of particles lies on the surface of a linear crystal represented by a periodic potential $V(x) = k/4\pi^2 cos(2\pi x)$.

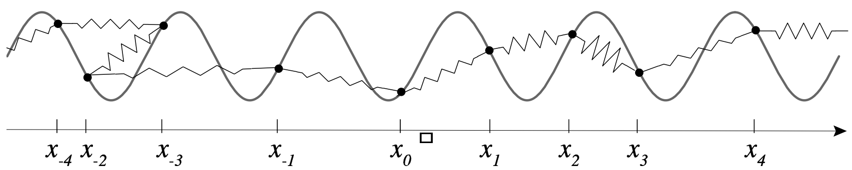


Fig. 6. 4. The Frenkel–Kontorova Model.

If x_k represents the location of the kth particle of the chain, this particle is in equilibrium whenever the sum of the forces applied to it is null:

$$(x_{k+1} - x_k) - (x_k - x_{k-1}) - \frac{k}{2\pi} sin(2\pi x_k) = 0 \tag{6.1}$$

This equation can be rewritten $dW = 0$ where W, the energy of the configuration of particles is given by :

$$W = \sum_k S(x_k, x_{k+1}) = \sum_k \frac{1}{2}(x_k - x_{k+1})^2 + \frac{k}{4\pi^2} cos(2\pi x_k).$$

We recognize S as the generating function of the Standard map. Hence *equilibrium states of the Frenkel-Kontorova model are in 1-1 correspondence with orbits of the Standard map.*

D. Billiard Maps

We revisit here the example of the billiard map presented in the introduction. Consider the dynamics of a ball in a convex, planar billiard. This ball is subject to simple laws : it goes in straight lines between two rebounds and the incidence and reflexion angles are equal at a rebound. We reproduce here a figure of the introduction:

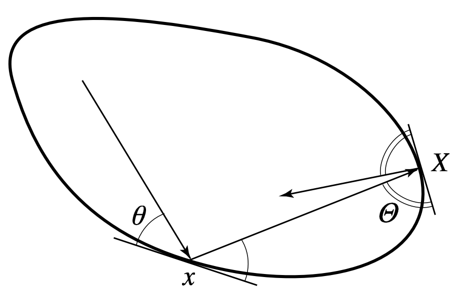


Fig. 6. 5. In a convex billiard, the point x and angle θ at a rebound uniquely and continuously determines the next point X and incidence angle Θ.

Let x be the arc length coordinate with respect to a given point on the boundary C of the billiard, which we orient counterclockwise. Let $y = -cos(\theta)$ where θ is the reflexion angle of a point of rebound. Because of the convexity of the billiard and the law of reflexion, a pair (x, y) at a rebound determines its successor (X, Y), and vice versa. Hence we have constructed a homeomorphism $f : (x, y) \mapsto (X, Y)$ of the (open) annulus $\mathbb{S}^1 \times (-1, 1)$ which is actually a C^{k-1} diffeomorphism if the boundary is C^k (LeCalvez (1990)). We call f the *billiard map*. If we increase y while keeping x fixed, the convexity of C implies that $C(X)$ moves in the positive direction along C. Thus:

$$\frac{\partial X}{\partial y} > 0 \tag{6.2}$$

and the billiard map satisfies the positive twist condition.

We now show that f is exact symplectic by exhibiting a generating function for it. Let $S(x, X) = \|C(X) - C(x)\|$ then, since $C' = \frac{dC}{dx}$ is a unit tangent vector:

$$\begin{aligned} \frac{\partial S}{\partial x} &= \frac{1}{S(x,X)}\left[C'(x).(C(x) - C(X))\right] = y \\ \frac{\partial S}{\partial X} &= \frac{-1}{S(x,X)}\left[C'(X).(C(X) - C(x))\right] = -Y \end{aligned} \tag{6.3}$$

which is to say:

$$YdX - ydx = -dS(x, X) \tag{6.4}$$

Thus, for the billiard map, the action function $W = \sum S(x_k, x_{k+1})$ is nothing more than the perimeter of the trajectory segment considered. For instance, periodic trajectories correspond to polygons in a given p, q-type who are critical points for the perimeter function. Figure 6. 6 shows that a period 5 orbit might come in different orders.

Exercise 6.2 Show that the billiard map for the round billiard is given by:

$$f(x, y) = (x + 2cos^{-1}(-y), y).$$

Exercise 6.3 Show that, for the billiard map, the equation $dW = 0$ expresses the equality between the angle of incidence and the angle of reflexion at each rebound.

7. The Poincaré-Birkhoff Theorem

In this section, we give a complete proof of the Poincaré-Birkhoff theorem, also called Poincaré's last theorem. We refer to Section 3.0for some motivation for this theorem. We use here some elementary material about circle diffeomorphisms, which the reader can get familiarized with in the appendix at the end of Chapter AM. We also use techniques of Conley for the gradient flow of the action function that the reader can read about in Appendix 2 or TOPO. We consider a map f of the compact annulus $\mathbf{A} = \mathbb{S}^1 \times [0, 1]$ and its lift F to $\mathcal{A} = \mathbb{R} \times [0, 1]$. We do *not* assume that f is a twist map, but rather that the restriction of F to each boundary component $u_\pm$, which are lifts of circle diffeomorphisms, have rotation numbers $\rho_\pm$ of $F\big|_{u_\pm}$ which satisfy $\rho_- < \rho_+$ (See . We say that F satisfies the *boundary twist condition.*

Theorem 7.1 (Poincaré-Birkhoff) *The lift F of an area preserving map of $\mathcal{A}$ which satisfies the boundary twist condition with $\rho_- < 0 < \rho_+$ has at least two fixed points. More generally, if $m/n \in [\rho_-, \rho_+]$, and m, n are coprime then F has at least two m, n–orbits.*

Proof. We follow the proof of LeCalvez (astérisque) , which is based on the following simple lemma:

Lemma (Decomposition) 7.2 *Any area preserving map f of a bounded annulus $\mathbf{A}$ isotopic to the Identity, can be written as a composition of twist maps:*

$$f = f_{2K} \circ \ldots \circ f_1$$

Proof. It is a general fact (left as an exercise to the reader) about topological groups that the connected component of the neutral element is generated by finite products of elements in any given neighborhood U of the neutral element of the group. Let f_0 be the shear map $f(x, y) = (x + y \mod 1, y)$. Since the set of maps satisfying the twist condition is open, there is a neighborhood U of Id in the set of area preserving maps of $\mathbf{A}$ which is such that $f \in U \Rightarrow f_0^{-1} \circ f$ is a negative twist map . Hence any f in U can be written as: $f = f_0 \circ (f_0^{-1} \circ f)$, a composition of two twist maps (one positive, the other negative). The group of area and orientation preserving maps of the annulus being connected, any map in that group can be written as a finite combinations of f as above. □

Let f be area preserving and let F be a lift of f to the covering space $\mathcal{A}$. Then $F = F_{2K} \circ \ldots \circ F_1$ where F_k lifts a twist map f_k. Let S_k be the generating function for F_k. If we let

$$W_0(\boldsymbol{x}) = \sum_{k=1}^{2K} S_k(x_k, x_{k+1}) \qquad \boldsymbol{x} \in X_{0,2K} = \{x_{2K+l} = x_l\}$$

then the Critical Action Lemma 5.3 shows that the critical points of W_0 correspond to periodic orbits under the successive f_k's, and hence to fixed points of f. To find these critical points, we study the gradient flow ζ^t of $-W_0$ and exhibit a compact set P of $X_{0,2K}$ which must contain critical points for the action. The set P is an *isolating block* in the sense of Conley, *i.e.* a compact neighborhood whose boundary points immediately exit P in (small) positive or negative time (see Appendix 2 or TOPO). This condition on the boundary implies that the maximum invariant set for ζ^t is in the interior of P (hence the term "isolating").

Lemma 7.3 *Whenever $\rho_- < 0 < \rho_+$, the set*

$$P = \{\boldsymbol{x} \in X_{0,2K} \mid 0 \leq -\partial_1 S_k(x_k, x_{k+1}) \leq 1\}$$

is an isolating block for the gradient flow ζ^t of $-W_0$. Moreover,

$$P \simeq \mathbb{S}^1 \times [0, 1]^K \times [0, 1]^{K-1}$$

with exit set $P^- = \mathbb{S}^1 \times [0, 1]^K \times \partial([0, 1]^{K-1})$

Proof. Setting $y_k = -\partial_1 S_k(x_k, x_{k+1})$, the faces of the boundary ∂P of P can be written as $\{y_k = 0\}$ or $\{y_k = 1\}$ for $k \in \{1, \ldots, 2K\}$. The behavior of the flow at a face $y_k = 1$, say, is given by the sign of $\frac{dy_k}{dt} = \dot{y}_k$:

$$\dot{y}_k = -\frac{d}{dt}\left(\partial_1 S_k(x_k, x_{k+1})\right) = -\partial_{11} S_k(x_k, x_{k+1})\dot{x}_k - \partial_{12} S_k(x_k, x_{k+1})\dot{x}_{k+1} \tag{7.1}$$

We let $Y_k = \partial_2 S_k(x_k, x_{k+1})$, *i.e.* $F_k(x_k, y_k) = (x_{k+1}, Y_k)$. With this notation $-\frac{\partial W_0}{\partial x_k} = Y_{k-1} + y_k$, and Equation (7.0) reads:

$$\dot{y}_k = \partial_{11} S_k(x_k, x_{k+1})(Y_{k-1} - y_k) + \partial_{12} S_k(x_k, x_{k+1})(Y_k - y_{k+1}) \tag{7.2}$$

and the invariance of the boundary component $\mathbb{R} \times \{1\}$ of $\mathbb{R} \times [0,1]$ under F_k tells us that, when $y_k = 1$ then $Y_k = 1$ as well. Since $y_{k\pm1} \leq 1$ and hence $Y_{k-1} \leq 1$,

$$Y_{k-1} - y_k \leq 0, \quad Y_k - y_{k+1} \geq 0. \tag{7.3}$$

Assume that k is even. Then f_k is a positive twist map and $-\partial_{12} S_k(x_k, x_{k+1}) > 0$. We need to determine the sign of $\partial_{11} S(x_k, x_{k+1})$ on the subset $\{y_k = 1\}$ of ∂P. On this set, we have $x_k = a(x_{k+1})$ where a is the restriction of F_k^{-1} to $y = 1$, this latter set being parameterized by x. Since a is the lift of an orientation preserving circle diffeomorphism, we have $a'(x) > 0$ for all x. We differentiate the equation $1 = \partial S(a(x), x)$ with respect to x :

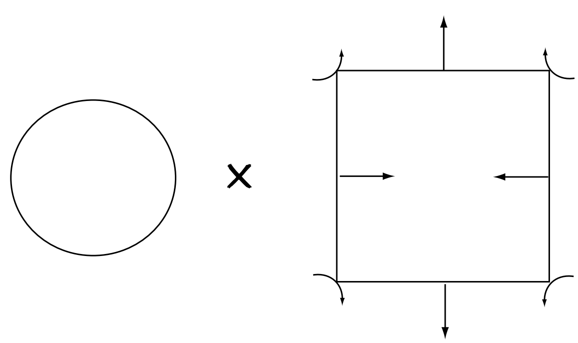
$$0 = a'(x)\partial_{11} S(a(x), x) + \partial_{12} S(a(x), x)$$

from which we deduce that $\partial_{11} S(x, a(x)) > 0$. Going back to Equation (7.2) , we see that if we are away from the boundary of the face $y_k = 1$ (i.e., in particular, $y_l \neq 1,\ l = k-1, k+1$), then the inequalities in (7.3) are strict, and we get $\dot{y}_k < 0$: the flow is strictly entering P through this face, or exiting it in negative time.

If we are on an edge of the face $y_k = 1$, the inequalities (7.3) may be equalities. But this cannot be the case for all k: if it were, $(x_k)_{k\in\mathbb{Z}}$ would be critical and (x_k, y_k) would be a fixed point for f on the boundary, which is impossible since then the rotation number $\rho_+ = 0$, a contradiction to $\rho_- < 0 < \rho_+$. So we can assume, say $Y_{l-1} - y_l < 0$, $y_l = y_{l+1} = \ldots = y_k = 1$, in which case (7.2) tells us that $\dot{y}_l \neq 0$ and the flow exits P in either positive or negative time at this point of ∂P.

The proof of the case k odd is exactly similar. We let the reader show in Exercise 7.4 that P and its exit set P^- have the topology advertised. □

This Lemma puts us in a situation which, since the work of Conley & Zehnder (1983) is a classic one in the field of symplectic topology. It can be schematized by the following diagram:



$S^1 \quad \times \quad [0,1]^K \times [0,1]^{K-1}$

Fig. 7. 2. The gradient flow at the boundary of the isolating block P

Given this topological data on its gradient flow, Proposition 50.3 tells us that W_0 must have at least $cl(\$^1) = 2$ of critical points. This completes the proof of the Poincaré-Birkhoff Theorem. The more general case of periodic orbits with rotation number $m/n \in (\rho_-, \rho_+)$ derives from the fixed point case by considering the map $F^n(\cdot) - (m, 0)$, which has new rotation numbers on the boundary $n(\rho_- - m/n) < 0 < n(\rho_+ - m/n)$ and whose fixed points correspond to m, n periodic orbits of F. □

Corollary CORvarprin or TWISTcorvarprin is 5.2, Proposition TWISTpropcritperiod is 5.3, Section TWISTsectionvariation is 5, Section TWISTsecexamples is 23,Section TWISTsecpb is 26

2 or AM

THE AUBRY-MATHER THEOREM

June 15 1999

I have decided to make a compact chapter, with no flow proof. That will come in the new GCchapter.

Action to be taken: Find the proof for the lemma extending annulus maps to cylinder maps. Find the drawing of Chenciner for Lipshitz condition on AM sets. Draw the 4 figures. Find the right reference for the no crossing lemma in refmanemml. Add a statement of KAM before AM. Proofread

8. Introduction

The orbits of the twist map f_0 whose lift is the completely integrable shear map given by $F_0(x, y) = (x+y, y)$, possess the following four fundamental properties, some of which we have yet to define:

(1) They lie on invariant circles which are graphs over the circle $\{y = 0\}$.

(2) They are ordered cyclically, as orbits of rotations on the circle.

(3) They come with all rotation numbers in $(-\infty, +\infty)$.

(4) They are action minimizers.

The KAM theorem (see THMkam) implies that, in the measure sense, most of these invariant circles will "survive" a *small* perturbation of f_0. The rotation number of these survivors has to be very irrational (diophantine). One cannot hope for all these circles to survive under arbitrary perturbation of the map f_0. In fact, it is known (ref ???: check jdm) that for $k > 0.9716354$, the standard map has no invariant circle. In the context of the Standard family, the Aubry-Mather theorem implies that, for each invariant circle of f_0, and for each $\lambda > 0$, there exists an invariant set for f_λ which can be seen as the remnant of the invariant circle. The properties of the orbits exhibited by the Aubry-Mather theorem will all be defined in subsequent sections.

Theorem 8.1 (Aubry-Mather) *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the lift of a C^2 twist map of the cylinder with generating function S satisfying the following growth condition:*

$$\lim_{|X-x|\to\infty} S(x, X) \to +\infty$$

Then F has orbits of all rotation numbers in $\mathbb{R}$ Moreover, these orbits can be chosen to have the following properties:

(1) They are cyclically ordered

(2) They lie on closed F-invariant sets, called Aubry- Mather sets *that form graphs over their projection on the circle $\{y = 0\}$ and are conjugated to closed invariant sets of lifts of circle homeomorphisms: either lifts of periodic orbits, Denjoy Cantor sets (and optionally, orbits homoclinic to these sets) or the full circle.*

(3) They may be chosen to be action minimizers.

We will see that an invariant Cantor sets must occur each time there is no invariant circle of a given irrational rotation number. The existence of these invariant Cantor sets was the striking novelty of this theorem. Often, the term "Aubry-Mather sets" is restricted to denote only the invariant Cantor sets.

Sketch of the proof

We will find periodic orbits of all rational rotation numbers by minimizing the periodic action W_{mn}. Aubry's Fundamental Lemma will imply that W_{mn}-minimizers are "cyclically ordered", *i.e.* ordered like orbits of circle homeomorphisms. The cyclic order property enables us to take limits of these periodic orbits (they will be in a compact set if their rotation numbers are in a bounded set). Cyclic order also implies that the rotation number of the limiting orbit exists and is the limit of the rotation numbers of the periodic orbits.

One way in which this presentation differs from the excellent surveys of this subject by Meiss (1992) or Hasselblat & Katok (1995) is the focus on the cyclic order property at the level of sequences (that are not necessarily realized by orbits). I found it a convenient bridge between the study of the dynamics of circle homeomorphisms (which appears in the appendix to this chapter) and that of Aubry-Mather sets.

We preceed our study by a Lemma, which implies that we can reduce our study to twist maps of the cylinder.

Lemma 8.2 *Let f be a $C^k, k \geq 2$, twist map of a compact annulus $\mathcal{A}$. Then f can be extended to a C^k twist map of the cylinder $\mathcal{C}$, in such a way that it coincides with the shear map $(x, y) \mapsto (x + cy, y)$ outside a compact set. In particular, letting $k \geq 2$ the generating function of any lift of the extended map satisfies the growth condition $\lim_{|X-x| \to \infty} S(x, X) \to +\infty$.*

As a corollary of this lemma, we can specialize the Aubry-Mather theorem to maps of the compact annulus:

Theorem 8.3 (Aubry-Mather on the compact annulus) *Let F be the lift of a twist map of the bounded annulus and suppose that the rotation numbers of the restriction of F to the lower and upper boundaries are ρ_-, and ρ_+ respectively. Then F has orbits of all rotation numbers in $[\rho_-, \rho_+]$. These orbits are minimizers, recurrent, cyclically ordered and they lie on compact invariant sets that form (uniformly) Lipshitz graphs over their projections. These sets may either be periodic orbits, invariant circles or invariant Cantor sets on which the map is semi-conjugate to lifts of circle rotations.*

Proof. ???

9. Cyclically Ordered Sequences and Orbits

If a map $G : \mathbb{R} \to \mathbb{R}$ is the lift of a circle homeomorphism which preserves the orientation, it is necessarily strictly increasing and must satisfy $G(x + 1) = G(x) + 1$. Hence, if $\{x_k\}_{k \in \mathbb{Z}}$ is an orbit of G, it must satisfy:

$$x_k \leq x_j + p \Rightarrow x_{k+1} \leq x_{j+1} + p, \forall k, j, p \in \mathbb{Z}. \tag{9.1}$$

We will say that a sequence $\{x_k\}_{k \in \mathbb{Z}}$ in $\mathbb{R}^{\mathbb{Z}}$ is *Cyclically Ordered*, (or *CO* in short) if it satisfies (9.1) . Clearly the CO sequences form a closed set for the *topology of pointwise convergence* in $\mathbb{R}^{\mathbb{Z}}$: $\boldsymbol{x}^{(j)} \to \boldsymbol{x}$ whenever $x_k^j \to x_k$ for all k. Not that this topology is the same as the product topology on the space of sequences. Using the *partial order on sequences*

$$\boldsymbol{x} < \boldsymbol{y} \Leftrightarrow \{x_k \leq y_k, \boldsymbol{x} \neq \boldsymbol{y}\},$$

we let the reader check that an equivalent definition of CO sequences is:

$$\forall m, n \in \mathbb{Z}, \qquad \tau_{m,n}\boldsymbol{x} \geq \boldsymbol{x} \quad \text{or} \quad \tau_{m,n}\boldsymbol{x} \leq \boldsymbol{x} \tag{9.2}$$

where

$$(\tau_{m,n}\boldsymbol{x})_k = x_{k+m} + n.$$

We will investigate this order relation and the maps $\tau_{m,n}$ in greater detail in GCchapter. We say that the orbit $\{(x_k, y_k)\}_{k \in \mathbb{Z}}$ of a twist map is a *Cyclically Ordered orbit* or *CO orbit* if $\{x_k\}_{k \in \mathbb{Z}}$ is CO. These orbits come with various other names in the litterature: *Well Ordered* (has no hint of the cyclic ordering), *Monotone* (is used in too many contexts), *Birkhoff* (this order was apparently never mentioned by Birkhoff) (3)

(3) This is not an indictment of the authors who have used these terminologies: the author of this book has himself used them all in various publications...

Lemma 9.1 *Let $\{x_k\}_{k\in\mathbb{Z}}$ be a CO sequence then $\rho(\boldsymbol{x}) = \lim_{k\to\infty} x_k/k$ exists and:*

$$|x_k - x_0 - k\rho(\boldsymbol{x})| \le 1. \tag{9.3}$$

Moreover $\boldsymbol{x} \to \rho(\boldsymbol{x})$ is a continuous function on CO sequences, when the set of sequences has been given the topology of pointwise convergence.

Define:

$$CO_{[a,b]} = \{\boldsymbol{x} \in CO \mid \rho(\boldsymbol{x}) \in [a, b]\}.$$

The following lemma shows that it is easy to find limits of CO sequences, as long as their rotation numbers are bounded.

Lemma 9.2 *The sets $CO_{[a,b]}/\tau_{1,0}$ and $CO_{[a,b]} \cap \{\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}} \mid x_0 \in [0, 1]\}$ are compact for the topology of pointwise convergence.*

We give the (simple) proofs of both these lemmas in the appendix to this chapter. The fact, given by these lemmas, that the rotation number behaves well under limits of CO-sequences is one of the essential points in the theory of twist maps that does not generalize to higher dimensional maps: to our knowledge, there is no canonical definition of CO sequences in $\mathbb{R}^n, n \ge 2$ which ensures the existence of rotation vectors which behave well under limits.

There is a visual way to describe CO sequences, which we now come to. A sequence $\boldsymbol{x}$ in $\mathbb{R}^{\mathbb{Z}}$ is a function $\mathbb{Z} \to \mathbb{R}$. One can interpolate this function linearly to give a piecewise affine function $\mathbb{R} \to \mathbb{R}$ that we denote by $t \mapsto x_t$. The graph of this function is sometimes called the *Aubry diagram* of the sequence. We say that two sequences $\boldsymbol{x}$ and $\boldsymbol{w}$ *cross* if their corresponding Aubry diagrams cross. There are two types of crossing: at an integer k, in which case $(x_{k-1} - w_{k-1})(x_{k+1} - w_{k+1}) < 0$ or at a non integer $t \in (k, k+1)$, in which case $(x_k - w_k)(x_{k+1} - w_{k+1}) < 0$. These inequalities can be taken as a definition of crossings. Non–crossing of two sequences can be put in terms of the strict partial order on sequence: $\boldsymbol{x}, \boldsymbol{y}$ do not cross if and only if $\boldsymbol{x} < \boldsymbol{y}$. *In particular a sequence $\boldsymbol{x}$ is CO if and only if it has no crossing with any of its translates $\tau_{m,n}\boldsymbol{x}$.*

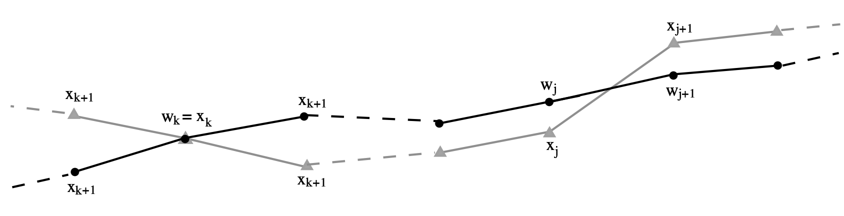


Fig. 9. 0. Aubry diagrams of sequences and their crossings: in this example the sequences $\boldsymbol{x}$ and $\boldsymbol{w}$ have crossings at the integer k and between the integers j and $j+1$.

10. Minimizing Orbits

A sequence segment $(x_k, \ldots, x_m)$ is *(action) minimizing* if

$$W(x_k, \ldots, x_m) \le W(y_k, \ldots, y_m)$$

for any other sequence segment $(y_k, \ldots, y_m)$ with same endpoints: $x_k = y_k, x_m = y_m$. Since minimizing segments are necessarily critical for W, they correspond to orbit segments called *(action) minimizing orbit segment* . A bi-infinite sequence is called a *(global action) minimizer* if any of its segments is minimizing and the orbit it corresponds to is a *minimizing orbit*, also called minimizer, when the context is clear. Note that the set of minimizers is closed under the topology of pointwise limit. Finally a W_{mn}*-minimizers* in X_{mm}, is a periodic sequences that minimize the function W_{mn}.

A recurrent theme in the Calculus of Variation is that minimizers have regimented crossings. In the case of geodesics on a Riemmanian manifold, geodesics that (locally) minimize length cannot have conjugate points, *i.e.* small variations with fixed endpoints of a minimizing geodesic only intersect that geodesic at the endpoints, (Milnor (1969)), and geodesics that minimize length globally cannot have self intersections (Manẽ (1991)). We will see, in the present theory, that minimizers satisfy a non-crossing condition, which implies that W_{mn}–minimizers are CO (and more generally that recurrent minimizers are CO).

Lemma 10.1 (crossing) *Suppose that* $(x - w)(X - W) \le 0$. *Then:*

$$S(x, X) + S(w, W) - S(x, W) - S(w, X) \le 0,$$

and equality occurs iff $(x - w)(X - W) = 0$

Proof. We can write:

$$S(x, X) - S(x, W) = \int_0^1 \partial_2 S(x, X_s)(X - W) ds,$$

where $X_s = (1 - s)W + sX$. Applying the same process to $h(x) = S(x, X) - S(x, W)$, we get:

$$S(x, X) + S(w, W) - S(x, W) - S(w, X) = h(x) - h(w) =$$
$$-\int_0^1 \int_0^1 \partial_{12} S(x_r, X_s)(X - W)(x - w) ds dr = \lambda(X - W)(x - w)$$

for some strictly negative λ, by the positive twist condition and for $x_r = (1 - r)w + rx$. □

The following is a watered down version of the Fundamental Lemma in Aubry & Le Daeron (1983). We follow Meiss (1992) :

Lemma 10.2 (Aubry's Fundamental Lemma) *Two distinct minimizers cross at most once.*

Proof. Suppose that $\boldsymbol{x}$ and $\boldsymbol{w}$ are two minimizers who cross twice. We perform some surgery on finite segments of $\boldsymbol{x}$ and $\boldsymbol{w}$ to get two new sequences $\boldsymbol{x}'$ and $\boldsymbol{w}'$ with at least one of them of lesser action, contradicting minimality. There are three cases to consider: (i) both crossings are at non integers, (ii) one crossing is at an integer, (iii) both crossings are at integers.

Case (i): Let $t_1 \in (i-1, i)$ and $t_2 \in (j, j+1)$ be the crossing times. Define:

$$x'_k = \begin{cases} w_k & \text{if} \quad k \in [i,j] \\ x_k & \text{otherwise} \end{cases} \qquad w'_k = \begin{cases} x_k & \text{if} \quad k \in [i,j] \\ w_k & \text{otherwise} \end{cases}$$

Letting W denote the action over an interval $[N, M]$ containing $[j-1, k+1]$, we easily compute that:

$$\begin{aligned} W(\boldsymbol{x}') + W(\boldsymbol{w}') - W(\boldsymbol{x}) - W(\boldsymbol{w}) = & \\ & S(x_{i-1}, w_i) + S(w_{i-1}, x_i) - S(x_{i-1}, x_i) - S(w_{i-1}, w_i) \\ & + S(x_j, w_{j+1}) + S(w_j, x_{j+1}) - S(x_j, x_{j+1}) - S(w_j, w_{j+1}). \end{aligned}$$

The Crossing Lemma 10.1shows that this difference of actions is negative, contradicting the minimality of $\boldsymbol{x}$ and $\boldsymbol{w}$.

Case (ii): In this case, only one crossing will contribute negatively to the difference of action of new and old sequences. We still get a contradiction.

Case (iii) Let $i-1$ and $j+1$ be the crossing times $\boldsymbol{x}$ amd $\boldsymbol{w}$, and construct $\boldsymbol{x}'$ and $\boldsymbol{w}'$ as before. In this case the difference in action between old and new segments is null. All the sequences must be minimizing, and hence correspond to orbits. But we have $x_{i-2} = w'_{i-2}, \quad x_{i-1} = w'_{i-1}$. Hence the points $\psi^{-1}(x_{i-2}, x_{i-1})$ and $\psi^{-1}(w'_{i-2}, w'_{i-1})$ of $\mathbb{R}^2$ are the same and generate a unique orbit under F. This in turn implies that $\boldsymbol{x} = \boldsymbol{w}$ are not distinct. □

Corollary 10.3 *W_{mn}-minimizing sequences are CO and their set is completely ordered for the partial order on sequences.*

Proof. Since the proof of Aubry's Lemma dealt with finite segments of sequences only, it also applies to show that two W_{mn}-*minimizers* in X_{mm}, may not cross twice within one period n. But two m, n-periodic sequences that cross once must necessarily cross twice within one period. Hence two W_{mn}-minimizers cannot cross at all. If $\boldsymbol{x}$ is a W_{mn} minimizer, $\tau_{i,j}\boldsymbol{x}$ is also a W_{mn}-minimizer. Since they do not cross, one must have either $\boldsymbol{x} < \tau_{i,j}\boldsymbol{x}$ or $\tau_{i,j}\boldsymbol{x} < \boldsymbol{x}$, for all $i, j \in \mathbb{Z}$, *i.e.* $\boldsymbol{x}$ is CO. □

We end this section by a proposition which we will need only in GCchapter.

Proposition 10.4 *Any W_{mn}-minimizer is a minimizer.*

Proof. We show that if $\boldsymbol{x}$ is a W_{mn}-minimizer is also a W_{kmkn} minimizer for any k. This implies that $\boldsymbol{x}$ is a minimizer on segments of arbitrary length: if $\boldsymbol{x}$ is a W_{kmkn} minimizer, any segment of $\boldsymbol{x}$ of length less than kn is minimizing. Hence $\boldsymbol{x}$ is a minimizer. Take a W_{kmkn}-minimizer $\boldsymbol{w}$. If $\boldsymbol{w}$ is not m, n-periodic, then $\boldsymbol{w}$ and $\tau_{m,n}\boldsymbol{w}$ are distinct. By the Corollary 10.3, they cannot cross. Suppose, say, that $\tau_{m,n}\boldsymbol{w} > \boldsymbol{w}$. Since $\tau_{m,n}$ trivially preserves the order on sequences, we must also have $\tau_{m,n}^k\boldsymbol{w} > \boldsymbol{w}$, a contradiction to the fact that $\boldsymbol{w}$ is km, kn- periodic. Hence $\boldsymbol{w}$ is in X_{mn} and its action over intervals of any length multiple of n cannot be less than that of $\boldsymbol{x}$. Hence $\boldsymbol{x}$ is also a W_{kmkn} minimizer. □

Exercise 10.5 Show that a minimizer corresponding to a recurrent (not necessarily periodic) orbit of the twist map is CO. (Remember that the orbit z_k of a dynamical system is called *recurrent* if z_0 is the limit of a subsequence z_{k_j}. Equivalently, z_0 is in its own ω–limit set). More generally, show that the set of minimizers of rotation number ω is completely ordered. (*Hint.* Mimic the proof of Proposition 10.4 : if an appropriate inequality is not satisfied, there must be a crossing. By recurrence, there is another one, a contradiction to Aubry's Lemma).

11. CO Orbits of All Rotation Numbers

A. CO Periodic Orbits

We prove that the set of W_{mn}-minimizers is not empty. By Corollary 10.3 this will show the existence of CO orbits of all rational rotation numbers.

Proposition 11.1 *Let the twist condition for the lift of a twist map F be uniform:*

$$\frac{\partial X(x,y)}{\partial y} > a > 0 \quad \forall (x,y) \in \mathbb{R}^2.$$

Then W_{mn} is proper and bounded below, and hence has a minimum.

We remind the reader that $h : X \to \mathbb{R}$ is *proper function* if the inverse image of a compact set is compact. If $X = \mathbb{R}^n$, then this translates to: the inverse image of any bounded interval is bounded. If h is also bounded below, it must indeed attain the $\inf_{x\in\mathbb{R}^n} = \alpha$ for some x_0 since, for instance, $h^{-1}[\alpha-1, \alpha+1]$ is compact.

Proof of Proposition 11.1 It is an immediate consequence of the following lemma (see MacKay & al. (1989)):

Lemma 11.2 *There is a constant α, and two strictly positive constants β and γ such that :*

$$S(x,X) \geq \alpha - \beta\,|X-x| + \gamma\,|X-x|^2$$

Proof. We can write:

$$S(x,X) = S(x,x) + \int_0^1 \partial_2 S(x, X_s)(X-x)ds,$$

where $X_s = (1-s)x + sX$. Applying the same process to $\partial_2 S$, we get:

$$\begin{aligned} S(x,X) = S(x,x) + \int_0^1 \partial_2 S(X_s, X_s)(X-x)ds \\ - \int_0^1 ds \int_0^1 \partial_{12} S(X_r, X_s)(X-x)^2 dr \end{aligned}.$$

We can conclude the proof of the lemma by taking

$$\alpha = \min_{x\in\mathbb{R}} S(x,x), \quad \beta = \max_{x\in\mathbb{R}} |\partial_2 S(x,X)|$$

(which exist by periodicity of S) and $\gamma = a/2$. □

B. CO Orbits of Irrational Rotation Numbers

The existence of CO orbits of irrational rotation numbers is a simple consequence of the existence of CO periodic orbits: pick a sequence $\boldsymbol{x}^{(k)}$ of W_{m_k,n_k}–minimizers, with $m_k/n_k \to \omega$ as $k \to \infty$. By using appropriate translations of the type $\tau_{m,0}$ on $\boldsymbol{x}^{(k)}$ (which neither change their rotation numbers, nor the fact that they are minimizers) we can assume that $\boldsymbol{x}^{(k)} \in [0,1]$. The sequence m_k/n_k is bounded and hence, by Corollary 10.3 the sequences $\boldsymbol{x}^{(k)}$ are in $CO_{[a,b]} \cap \{\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}} \mid x_0 \in [0,1]\}$ for some $a, b \in \mathbb{R}$. Lemma 9.2 garantees the existence of a converging subsequence in $CO_{[a,b]}$ and Lemma 9.1 shows that the limit of this subsequence has rotation number ω. Finally, note that the periods $n_k \to \infty$ as $k \to \infty$. In particular, any finite segment of $\boldsymbol{x}$ is the limit of minimizing segments, hence minimizing itself. □

12. Aubry-Mather Sets

We have proven Part (1) and (3) of the Aubry-Mather theorem: existence of cyclically ordered, minimizing orbits of all rotation numbers. We now prove Part (2) of the Aubry-Mather theorem: the cyclically ordered orbits that we found in the previous section lie on Aubry-Mather sets, which now describe.

We say that a set M in $\mathbb{R}^2$ is F*-ordered* if, for z, z' in M,

$$\pi(z) < \pi(z') \Rightarrow \pi(F(z)) < \pi(F(z')),$$

where π is the x-projection. If moreover M is invariant by F and F^{-1}, then the sequences $\boldsymbol{x}, \boldsymbol{x}'$ of x–coordinates of z and z' satisfy $\boldsymbol{x} \prec \boldsymbol{x}'$. An example of F-ordered invariant set is the set of points in a CO orbit and all their integer translates (In fact, this is an alternative definition of CO orbits). Note that an invariant circle for the map which is a graph (we will see in INVchapterthat all invariant circles are graphs) is F-ordered. We now want to explore the properties of F-ordered invariant sets. Crucial to the properties of these sets is the following *ratchet phenomenon* (I owe this terminology to G.R. Hall), which is a somewhat quantitative expression of the twist condition:

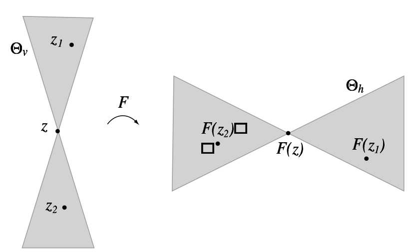


Fig. 12. 0. The ratchet phenomenon for the lift of a positive twist map F: there are two cones Θ_v and Θ_h in $\mathbb{R}^2$ centered around the y and x-axes respectively, such that, if z, z' are two points of $\mathbb{R}^2$ with $z' \in z + \Theta_v$, then $F(z') \in F(z) + \Theta_h$. More precisely, for a positive twist map $z' \in z + \Theta_v^+ \Rightarrow F(z') \in F(z) + \Theta_h^+$, where the half cones Θ_h^+, Θ_v^+ have the obvious meaning. The same holds for the half cones Θ_h^- and Θ_v^-. If g is *negative* twist (*eg.* F^{-1}), then the signs are reversed. The same cones can be used for F^{-1} as for F.

Lemma 12.1 *Let F be the lift of a twist map satisfying $\frac{\partial X}{\partial y} > a > 0$ in some region. Then, in that region, F satisfies the ratchet phenomenon for some cones Θ_v, Θ_h whose angles only depend on a.*

Proof. Left as an exercise.

Proposition 12.2 *The closure of an F–ordered invariant set is F–ordered and invariant.*

Proof. The invariance is by continuity of F. Suppose that, in the closure $\overline{M}$ of M there are z, z' in $\overline{M}$, with $\pi(z) < \pi(z')$ but $\pi(F(z)) = \pi(F(z'))$ (the worst scenario). By the ratchet phenomenon for F^{-1}, $F(z)$ must be above $F(z')$ and $\pi(F^2(z')) < \pi(F^2(z))$, *i.e.* the x orbits of z and z' switched order. This is impossible since in M the order is preserved.

Proposition 12.3 *If M is an F-ordered invariant set, then it is a Lipschitz graph over its projection: there exists a constant K depending only on F such that, if (x, y) and $(x\prime, y\prime)$ are two points of M, then:*

$$|y\prime - y| \leq K|x\prime - x|$$

with K only depending on the twist constant $a = \inf_M \frac{\partial X}{\partial y}$.

Note that a, and hence K can be chosen the same for all F-ordered sets in a compact region.

Remark 12.4 Applied to the special case of invariant circles, Proposition 12.3shows that any invariant circle for a twist map which is a graph is Lipschitz. This is a theorem originally due to Birkhoff, who also proved (see INVchapter) that all invariant circles for twist maps must be graphs.

Proof. The proof of Lemma 12.3 shows that if M is F-ordered, we cannot have z, z' in M and $\pi(z) = \pi(z')$ unless $z = z'$. Hence π is injective on M, and M is a graph. To show that M forms a *Lipschitz* graph over its projection, let z and z' be two points of M and $\boldsymbol{x}$ and $\boldsymbol{x}'$ the corresponding sequences of x-coordinates of their orbits. Assuming $\pi(z) < \pi(z')$, we must have $\boldsymbol{x} \prec \boldsymbol{x}'$. If $z' \in z + \Theta_v^+$, the ratchet phenomenon implies that $F^{-1}(z') \in F^{-1}(z) + \Theta_h^-$, *i.e.* $x'_{-1} > x_{-1}$, a contradiction. Likewise z' cannot be in the cone $z + \Theta_v^-$, and hence it must be in the cone complementary to Θ_v at z. This cone condition is easily transcribed into a uniform Lipschitz condition $|y' - y| < K|x' - x|$. □

Lemma 12.5 *All points in an F-ordered set have the same rotation number.*

Proof. This is a consequence of the fact (Lemma AMlemmax<yrot in the appendix) that if $\boldsymbol{x} < \boldsymbol{x}'$ are two CO sequences, they must have the same rotation number. □

Definition 12.6 An *Aubry-Mather set* M for the lift F of a twist map f of the cylinder is a closed, F-ordered set invariant under F, F^{-1} and the integer translation T.

Theorem 12.7 (Properties of Aubry-Mather sets) *Let M be an Aubry-Mather set for a lift F of a twist map of the cylinder.*

(a) M forms a graph over its projection $\pi(M)$, which is Lipschitz with Lipschitz constant only depending on a where $\frac{\partial X}{\partial y > a}$.

(b) All the orbits in M are cyclically ordered and they all have the same rotation number, which is called the rotation number of M.

(c) The projection $\pi(M)$ is a closed invariant set for the lift of a circle homeomorphism, and hence F restricted to M is conjugated to the lift of a circle homeomorphism via π.

We remind the reader that a *conjugacy* between two maps $F : M \to M$ and $G : N \to N$ is a homeomorphism $h : M \to N$ such that $h \circ F = H \circ h$. Taking the closure of all the integer translates of the points in the CO orbits found in the previous section, we immediately get:

Theorem 12.8 *Let F be the lift of a twist map of the cylinder. Then F has Aubry-Mather sets of all rotation numbers in $\mathbb{R}$. Any CO orbit is in an Aubry-Mather set.*

Note that this theorem gives part (b) of the Aubry-Mather theorem.

Proof of Theorem 12.7 We have shown in Lemmas 12.6 and 12.5 that (a) and (b) are in fact properties of invariant F-ordered sets. As for Property (c), since π is one to one on M, F induces a continuous (Lipschitz, in fact) increasing map G on $\pi(M)$ (by $G(\pi(z)) = \pi(F(z))$. Since M and thus $\pi(M)$ are invariant under integer translation, we have $G(x + 1) = G(x) + 1$. The set $\pi(M)$ is closed and invariant under integer translation since M is. If $\pi(M) = \mathbb{R}$ then G is the lift of a circle homeomorphism. If $\pi(M) \neq \mathbb{R}$, then its complement is made of open intervals. The fact that G is increasing on $\pi(M)$ allows one to extend G by linear interpolation on each interval in the complement of $\pi(M)$. The resulting map G is increasing, continuous and $G(x + 1) = G(x) + 1$, hence the lift of a circle homeomorphism. By construction $G(\pi(z)) = \pi(F(z))$, and $\pi\big|_M$ is a continuous, 1-1 map on the compact set M, hence a homeomorphism $M \to \pi(M)$. Thus π is a conjugacy between F on M and G on $\pi(M)$, which is a closed and invariant set under G and G^{-1}. □

If G is the lift of a circle homeomorphism constructed in the proof of Theorem 12.7, the possible dynamics for invariant sets of circle maps described in the appendix become, under the conjugacy, possible dynamics on Aubry-Mather sets M for F. Hence an Aubry-Mather set M is either:

(i) an ordered collection of periodic orbits with (possibly) heteroclinic orbits joining them, or

(ii) the lift of an f-invariant circle, or

(iii) an F-invariant Cantor set with (possibly) homoclinic orbits in its gaps.

The rotation number of M is necessarily rational in Case (i), and necessarily irrational in Case (iii). In Case (ii), M may have either rational or irrational rotation number, as the example of the shear map shows. However, it has been shown (Zehnder (???generic prop of twist maps)) that maps with rational invariant circles are non generic. As for homoclinic and heteroclinic orbits as in (i) and (iii), they have been shown to exist each time there are no invariant circles of the corresponding rotation numbers Hasselblat & Katok (1995) , Mather (1986) .

The feature that is striking in the Aubry-Mather Mather theorem is the possible occurence of Aubry-Mather sets as in (iii). The F-invariant Cantor sets have been called *Cantori* by Percival (1979)who constructed them for the discontinuous sawtooth map (a standard map with sawtooth shaped potential). This type of dynamics does occur in twist map, since it can be shown that many maps have no invariant circles, and hence the irrational Aubry-Mather sets must be of type (iii), *i.e.* contain a Cantori.

Although one can construct Aubry-Mather sets that are not made of minimizers, the name "Aubry Mather set" is often reserved to the action minimizing Cantori M_ω as defined below:

Proposition 12.9 *For each rotation number ω there is a unique Cantorus M_ω made of recurrent minimizing orbits of rotation number ω. The closure of any CO minimizing orbit of rotation number ω is contained in M_ω.*

Proof. A CO minimizing orbit forms an F-ordered set, contained in an Aubry-Mather set, and hence conjugated to an orbit of a circle homeomorphism. The closure of the CO minimizing orbit is therefore in a Cantorus, conjugated to the ω–limit set of the circle homeomorphism. As limit of minimizers, this Cantorus is made up of minimizers. We now prove that this Cantorus is unique: suppose not and there are two of them. Exercise 10.5 implies that the (disjoint) union of these two Cantori forms an F–ordered set, hence conjugated to a closed invariant set of a circle homeomorphism. Each Cantorus is the ω-limit set of its points. This is a contradiction to the uniqueness of ω limit sets of circle homeomorphisms proven in Theorem AMthmcircleomlimset. □

13.1 Appendix: Cyclically Ordered Sequences and Circle Maps

In this section, we prove Lemma 9.1, and Lemma 9.2. We then recover important facts about circle homeomorphisms and their invariant sets using the language of CO sequences. Part of the proof below is classical, due to Poincaré in his study of circle homeomorphisms.

A. Proof of Lemmas 9.1 and 9.2

Proof of Lemma 9.1. Let $\boldsymbol{x}$ be a CO sequence. We want to prove that the sequence $\{\frac{x_n - x_0}{n}\}_{n\in\mathbb{Z}}$ is a Cauchy sequence as $n \to \pm\infty$. We do the case $n \to +\infty$ here, the case $n \to -\infty$ will follow.

Given $n \in \mathbb{N}$, let α_n be the integer such that:

$$x_0 + \alpha_n \le x_n \le x_0 + \alpha_n + 1. \tag{13.1}$$

We prove by induction that

$$x_0 + k\alpha_n \le x_{kn} \le x_0 + k\alpha_n + k, \quad \forall k \in \mathbb{N}. \tag{13.2}$$

Indeed, step 1 in the induction is just (13.1) , and if we assume step k, *i.e.* (13.2) then, since $\boldsymbol{x}$ is CO, we get

$$x_n + k\alpha_n \le x_{(k+1)n} \le x_n + k\alpha_n + k.$$

Using (13.1) this gives $x_0 + (k+1)\alpha_n \le x_{(k+1)n} \le x_0 + (k+1)\alpha_n + (k+1)$, which is the step $k+1$ and finishes the induction.

Dividing (13.2) by k we get

$$\alpha_n \le \frac{x_{kn} - x_0}{k} \le \alpha_n + 1. \tag{13.3}$$

Since this is true for all $k > 0$,

$$\left| \frac{x_{kn} - x_0}{k} - \frac{x_n - x_0}{1} \right| \le 1 \Rightarrow \left| \frac{x_{kn} - x_0}{kn} - \frac{x_n - x_0}{n} \right| \le \frac{1}{|n|}. \tag{13.4}$$

Writing $z_n = \frac{x_n - x_0}{n}$, and assuming $m > 0, n > 0$ we have that

$$|z_n - z_m| \le |z_n - z_{mn}| + |z_{mn} - z_m| \le \frac{1}{n} + \frac{1}{m}, \tag{13.5}$$

and hence $z_n, n \in \mathbb{N}$, is a Cauchy sequence whose limit we call $\rho(\boldsymbol{x})$.

To see how the case $n \to -\infty$ follows, let $m \to \infty$ in (13.5) , and multiply by n:

$$|x_n - x_0 - n\rho(\boldsymbol{x})| \le 1. \tag{13.6}$$

Since in all the above we could have replaced x_0 by an arbitrary $x_m, m \in \mathbb{Z}$, the following also holds:

$$|x_n - x_m - (n - m)\rho(\boldsymbol{x})| \le 1 \qquad \forall m, n \in \mathbb{Z}. \tag{13.7}$$

We let the reader check that this last inequality implies that $\lim_{n \to -\infty} z_n = \rho(\boldsymbol{x})$.

The continuity of ρ is also a consequence of Formula (13.6) . Suppose $\boldsymbol{x}^{(j)} \to \boldsymbol{x}$ pointwise as $j \to \infty$. Constructing sequences $\boldsymbol{z}^{(j)}$ as above, and denoting $\rho(\boldsymbol{x}^{(j)}) = \omega_j, \rho(\boldsymbol{x}) = \omega$, (13.6) yields

$$|z_k^{(j)} - \omega_j| \le \frac{1}{k}, \quad |z_k - \omega| \le \frac{1}{k}. \tag{13.8}$$

Since $\boldsymbol{z}^{(j)} \to \boldsymbol{z}$, for all k and $\epsilon > 0$,

$$|\omega_j - \omega_i| \le |\omega_j - z_k^{(j)}| + |z_k^{(j)} - z_k^{(i)}| + |z_k^{(i)} - \omega_i| \le \frac{2}{k} + \epsilon$$

whenever i, j are big enough. Hence $\{\omega_k\}_{k \in \mathbb{Z}}$ is a Cauchy sequence, whose limit we denote by ω. Letting $j \to \infty$ in (13.8) yields $\omega = \rho(x)$. □

Proof of Lemma 9.2 Lemma 9.1 implies that $\mathrm{CO}_{[a,b]} \cap \{\boldsymbol{x} \mid x_0 \in [0,1]\}$ is a closed subset of the set:

$$\{\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}} \mid x_k = x_0 + k\omega + y_k, (x_0, \omega, \boldsymbol{y}) \in [0,1] \times [a,b] \times [-1,1]^{\mathbb{Z}}, \text{with } y_0 = 0\}$$

which is compact for the product topology, by Tychonov's theorem. We let the reader derive a similar proof for $\mathrm{CO}_{[a,b]}/\tau_{1,0}$. □

B. Dynamics of Circle Homeomorphisms

The orbits of an orientation preserving circle homeomorphism are (by definition!) Cyclically Ordered. From Lemma 9.1, we can deduce the following theorem, due to Poincaré (1985):

Theorem 13.1 *All the orbits of the lift F of an orientation preserving circle homeomorphism f have the same rotation number, denoted by $\rho(F)$. The rotation number ρ is a continuous function of F, where the set of lifts of homeomorphisms of the circle is given the C^0 topology.*

Proof. We start by a simple but useful lemma.

Lemma 13.2 *If two* CO *sequences* $\boldsymbol{x}, \boldsymbol{x}'$ *satisfy* $\boldsymbol{x} < \boldsymbol{x}'$ *then* $\rho(\boldsymbol{x}) = \rho(\boldsymbol{x}')$.

Proof. The rotation numbers are the respective asymptotic slopes of the Aubry diagram of $\boldsymbol{x}$ and $\boldsymbol{x}'$. If $\rho(\boldsymbol{x}) \neq \rho(\boldsymbol{x}')$, the the Aubry diagram must cross: there must be a k_0 and a k_1 such that $x_{k_1} > x'_{k_1}$ and $x_{k_1} < x'_{k_1}$. That contradicts $\boldsymbol{x} < \boldsymbol{x}'$. □

Continuing with the proof of Theorem 13.1, since F is increasing, two distinct orbits $\boldsymbol{x}$ and $\boldsymbol{w}$ of F satisfy $\boldsymbol{x} < \boldsymbol{w}$ or $\boldsymbol{w} < \boldsymbol{x}$. From the lemma $\boldsymbol{x}$ and $\boldsymbol{w}$ have same rotation number. If $f_n \to f$ in the C^0 topology, then the f_n orbit of a point x (a CO sequence) tends pointwise to the f orbit of x. By Lemma 9.1, $\lim \rho(f_n) = \lim \rho(\{f_n^k(x)\}) = \rho(\{f^k(x)\}) = \rho(f)$. □

We now remind the reader about the structure of invariant sets of circle homeomorphisms. Remember that the *Omega limit set* $\omega(x)$ of a point x under a dynamical system f on some space X is the set of limit points of all subsequences $\{x_{k_j}\}$ where $x_k = f^k(x)$ and $k_j \to +\infty$ as $j \to +\infty$, *i.e.* the set of limit points of the forward orbit. Likewise, the *Alpha limit set* $\alpha(x)$ is the set of limit points of the backward orbit. The following theorem, which basically appears in Poincaré (1985), classifies the possible dynamics of circle homeomorphisms:

Theorem 13.3 *Let f be a circle homeomorphism and F a lift of f. If $\rho(F)$ is rational, then, for any $x \in \mathbb{S}^1$, $\omega(x)$ and $\alpha(x)$ are periodic orbits. The orbit of x is either periodic (in which case $x \in \omega(x) = \alpha(x)$) or it is heteroclinic between $\alpha(x)$ and $\omega(x)$.*

If $\rho(F)$ is irrational, then, for any $x, x' \in \mathbb{S}^1$, $\alpha(x) = \alpha(x') = \omega(x) = \omega(x')$. Call this set $\Omega(f)$. Then $\Omega(f)$ is either the full circle, or a minimal invariant set which is a Cantor set. In the first case any orbit is dense in the circle, and f is conjugated to a rotation by $\rho(F)$. In the second case, either $x \in \Omega(f)$ is recurrent, or it is homoclimic to $\Omega(f)$, a "gap orbit".

We remind the reader that a *Cantor set* K is a closed, perfect, and totally disconnected topological set. *Perfect* means that each point in K is the limit of some (non constant) sequence in K, and *totally disconnected* means that, given any two points a and b in K, one can find disjoint closed sets A and B with $a \in A, b \in B$ and $A \cup B = K$. In the real line or the circle, a closed set is totally disconnect if and only if it is nowhere dense. A set X is *nowhere dense* if $Interior(Closure(X)) = \emptyset$.

Proof of Theorem 13.3

Rational rotation number. Suppose $\rho(F) = m/n$. Then $F^n - m$ must have a fixed point, otherwise for all $x \in \mathbb{R}$, $F^n(x) - x \neq m$ and we can assume $F^n(x) - x > m$. By compactness of $\mathbb{S}^1$, $\rho(F) > m/n$, a contradiction. Hence F has an m, n-periodic orbit. By continuity, on any interval I where $F^n - Id - m$ is non zero, it must stay of a constant sign. This sign describes the direction of progress of points inside the orbit of I towards its endpoints: they must be heteroclinic to the endpoint orbits. Conversely, if F has an m, n-periodic orbit, its rotation number and thus that of F must be m/n.

Irrational rotation number. Suppose $\rho(F)$ is irrational. Let $x \in \mathbb{S}^1$ and denote by $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{Z}}$ its orbit under f (with $x = x_0$). Suppose $\omega(x) = \mathbb{S}^1$. Then $\omega(x') = \mathbb{S}^1$ for any other $x' \in \mathbb{S}^1$, otherwise there would be an interval (a, b) not containing any $x'_k = f^k(x')$. But (a, b) would contain some $[x_n, x_m]$ by density of $\boldsymbol{x}$. The intervals $f^{-i(m-n)}[x_n, x_m]$ must cover $\mathbb{S}^1$ and hence $f^{i(m-n)}x' \in (a, b)$ for some i, a contradiction. We guide the reader through the proof that f is conjugated to a rotation by $\rho(f)$ in Exercise 13.5.

Suppose $\omega(x) \neq \mathbb{S}^1$. Then, since $\omega(x)$ is closed, its complement is the union of open intervals. Take another point x'. We want to show that $\omega(x') = \omega(x)$. We will prove that $\omega(x') \subset \omega(x)$: by symmetry $\omega(x) \subset \omega(x')$. This is obvious if $x' \in \omega(x)$. Suppose not. Then x' is in an open interval I in the complement of $\omega(x)$ whose endpoints are in $\omega(x)$. The orbit of I is made of open intervals in the complement of $\omega(x)$ whose endpoints are orbits in $\omega(x)$. Since there is no periodic orbit, these intervals are disjoint: by the intermediate value theorem $f^k(I) \subset I$ would imply the existence of a fixed point for f^k, hence a periodic orbit. The length of these intervals must tend toward 0 under iteration. Thus the orbit of x' approaches the endpoint orbit of I arbitrarily *i.e.* it is asymptotic to $\omega(x)$. Hence $\omega(x') \subset \omega(x)$. In particular $\omega(x) = \Omega(f)$ is a minimal invariant set: any closed invariant subset of $\Omega(f)$ must contain the ω-limit set of any of its point, hence $\Omega(f)$ itself.

We now show that $\Omega(f)$ is a Cantor set. That it is closed is a property of ω-limit sets. It is perfect since $x \in \Omega(f)$ means that $x \in \omega(x)$ and hence $f^{n_k}(x) \to x$ for some $n_k \nearrow \infty$ with all the $f^{n_k}(x)$'s are in $\omega(x)$. To prove that $\Omega(f)$ is nowhere dense, first note that the topological boundary $\partial\Omega(f) = \Omega \backslash Interior(\Omega(f))$ must satisfy $\partial\Omega(f) = \Omega(f)$ or $\partial\Omega(f) = \emptyset$: $\partial\Omega(f)$ is closed, invariant under f and included in $\Omega(f)$ which is a minimal set. But $\partial\Omega(f) = \emptyset$ means $\Omega(f) = Interior(\Omega(f))$ is open, and because it is also closed, it must be all of $\mathbb{S}^1$, which we have ruled out. The alternative is $\partial\Omega(f) = \Omega(f)$, which means $Interior(\Omega(f)) = \emptyset$, what we wanted to prove. □

Remark 13.4 A circle homeomorphism with an invariant Cantor set cannot be too smooth: Denjoy (see Hasselblat & Katok (1995) , Robinson (1994)) proved that if f is a C^1 diffeomorphism of $\mathbb{S}^1$ with irrational rotation number and derivative of bounded variation, then f has a dense orbit (*i.e.* $\Omega(f) = \mathbb{S}^1$) and is therefore conjugated to a rotation of angle $\rho(F)$. On the other hand, Denjoy did construct a C^1 diffeomorphism with $\Omega(f)$ a Cantor set. The idea is simple: take a rotation by irrational angle α. Cut the circle at some point x and at all its iterate $f^k(x)$. Glue in at these cuts intervals I_k of length going to 0 as $k \to \infty$, in such a way that the new space you obtain is again a circle. Extend the map f by linear interpolation on the I_k. You get a circle homeomorphism with rotation number α. With some care, one can make this homeomorphism differentiable, but only up to a point (C^1 with Hölder derivative). The complement of the I_k's in the new circle is a Cantor set, which is minimal.

Exercise 13.5 In this exercise, we prove that if a circle homeomorphism has a dense orbit, then it is conjugated to a rotation.
a) Prove that $\boldsymbol{x}$ is a CO sequence with irrational $\rho(\boldsymbol{x})$ *iff*

$$\forall n, m, p \in \mathbb{Z}, \quad x_n < x_m + p \Longleftrightarrow n\rho(\boldsymbol{x}) < m\rho(\boldsymbol{x}) + p$$

(*Hint.* Use Formula (13.7) for multiples of m and n). What is the proper corresponding statement for CO sequences of rational rotation number?
b) Suppose the circle homeomorphism f has a dense orbit $\boldsymbol{x}$. Build a map $h : \mathbb{S}^1 \to \mathbb{S}^1$ by first defining it on $\boldsymbol{x}$ by:

$$x_k \mapsto k\rho(\boldsymbol{x})$$

Use a) to show that h is order preserving and show that its extension by continuity is well defined, has continuous inverse and preserves orbits.

Lemma AMlemmax<yrot is 13.2, Theorem AMtheoremperiodic is 6.3, Section AMsectionlimits is 7, Lemma AMlemmaaubry is 10.2, Corollary AMcorollaryaubry is 10.3, Exercise AMexominordered is 10.5, Lemma AMlemmacoestimate is 9.1, Lemma 11.2is AMthmconvest, Proposition AMpropwmnmin is 10.4, Proposition AMproplipschitz is AMthmlipschitzProposition AMpropmom is 12.9, Theorem AMthmcircleomlimset is 13.3, Formula AMformqgeod is (13.6)

CHAPTER 3 or GC

GHOST CIRCLES

June 30, 1999

In Chapter AM, we saw how traces of the invariant circles of the completely integrable map persist, sometime in the weak form of invariant Cantor sets, in any twist map. The main result of this chapter, Theorem GCthmamordered, provides a vertical ordering of these Aubry-Mather sets in the cylinder, by showing that they belong to family of *nontrivial ???* circles that are graph over the circle $\{y = 0\}$. These circles are mutually disjoint and are ordered according to the rotation number of the Aubry-Mather sets. This result was written in Angenent & Golé (1991).

To do this, we establish important properties of the gradient flow of the action functional in the space of sequences. The central property, given by the Sturmian Lemma, is that the intersection index of two sequences cannot increase under the gradient flow of the action. One important consequence is that the flow is monotone: it preserves the natural order between sequences. This fact yields a new proof of the Aubry-Mather Theorem. It also enables us to define special invariant sets for the gradient flow that we called ghost circles, which we study in some detail here. The family of circles that neatly arranges the Aubry-Mather sets are projections of ghost circles in the cylinder.

14. Gradient Flow of the Action

A. Definition of the Flow

We consider a twist map f of the cylinder and its lift F whose generating function S is C^2. For simplicity, we will assume that the second derivative of S is bounded. This assumption is satisfied for twist maps of the bounded annulus which are extended to maps of the cylinder as in AMlemmaextend. In this section we investigate the property of the "gradient" flow of the action associated to the generating function S of F solution to:

$$-\nabla W(\boldsymbol{x})_k = \dot{x}_k = -[\partial_1 S(x_k, x_{k+1}) + \partial_2 S(x_{k-1}, x_k)] \tag{14.1}$$

Since this is an infinite system of ODEs, we need to set up the proper spaces to talk about such a flow.

We endow $\mathbb{R}^{\mathbb{Z}}$ with the norm :

$$\|\boldsymbol{x}\| = \sum_{-\infty}^{+\infty} \frac{|x_k|}{2^{|k|}}$$

We let X be the subspace of $\mathbb{R}^{\mathbb{Z}}$ of elements of bounded norm, which is a Banach space. On bounded subsets of X, the topology given by the above norm is equivalent to the *product topology*, itself equivalent to the *topology of pointwise convergence*. Remember from Chapter AM that $\mathbb{Z}^2$ acts on $\mathbb{R}^{\mathbb{Z}}$ by:

$$(\tau_{m,n}\boldsymbol{x})_k = x_{k+m} + n$$

The map $\tau_{0,1}$ which we also denote by T has the effect of translating each term of the sequence by 1. The map $\tau_{1,0}$ which we denote also by σ is called the *shift map*, as it shifts the indices of a sequences by 1.

We define $X/\mathbb{Z} := X/\tau_{0,1}$ and we can choose as a representative of a sequence $\boldsymbol{x}$ one such that $x_0 \in [0, 1)$. More generally, in this chapter, the quotient of any subset of $\mathbb{R}^{\mathbb{Z}}$ by $\mathbb{Z}$ will be with respect to the action of the translation $T = \tau_{0,1}$.

Proposition 14.1 *Suppose that the generating function S is C^2 with bounded second derivative. The infinite system of O.D.E's*

$$-\nabla W(\boldsymbol{x})_k = \dot{x}_k = -[\partial_1 S(x_k, x_{k+1}) + \partial_2 S(x_{k-1}, x_k)] \tag{14.2}$$

defines a C^1 local flow ζ^t on X as well as on $X/\mathbb{Z}$, for the topology of pointwise convergence. The rest points of ζ^t on X correspond to orbits of the map F.

Proof. We prove that the vector field $-\nabla W$ is C^1 by exhibiting its differential. The proposition follows from general theorems on existence and uniqueness of solutions of ODEs in Banach spaces (*eg.* Lang (1983) , Theorems 3.1 and 4.3). The following map is the derivative of $\boldsymbol{x} \mapsto -\nabla W(\boldsymbol{x})$:

$$L : \{v_k\}_{k\in\mathbb{Z}} \mapsto \{\beta_k v_{k-1} + \alpha_k v_k + \beta_{k+1} v_{k+1}\}_{k\in\mathbb{Z}}$$
$$\alpha_k = -\partial_{22} S(x_{k-1}, x_k) - \partial_{11} S(x_k, x_{k+1}), \quad \beta_k = -\partial_{12} S(x_{k-1}, x_k)$$

Indeed, this map is linear with (uniformly) bounded coefficients, hence a continuous linear operator. Clearly:

$$-\nabla W(\boldsymbol{x})\boldsymbol{v} - L(\boldsymbol{v}) = \|\boldsymbol{v}\| \, \psi(\boldsymbol{v})$$

with $\lim_{\boldsymbol{v}\to 0} \psi(\boldsymbol{v}) = 0$. $\square$

B. Order Properties of the Flow

$\mathbb{R}^{\mathbb{Z}}$ is partially ordered by:

$$\boldsymbol{x} \leq \boldsymbol{y} \Leftrightarrow \forall_{k\in\mathbb{Z}} \quad x_k \leq y_k.$$

We also define $\boldsymbol{x} < \boldsymbol{y}$ to mean $\boldsymbol{x} \leq \boldsymbol{y}$, but $\boldsymbol{x} \neq \boldsymbol{y}$; and we write $x \prec y$ to denote the condition $x_j < y_j$ for all $j \in \mathbb{Z}$. The *order interval* $[\boldsymbol{x}, \boldsymbol{y}]$ is defined by:

$$[\boldsymbol{x}, \boldsymbol{y}] = \{\boldsymbol{z} \in \mathbb{R}^{\mathbb{Z}} \mid \boldsymbol{x} \leq \boldsymbol{z} \leq \boldsymbol{y}\}$$

and the *positive order cone* at $\boldsymbol{x}$

$$V_+(\boldsymbol{x}) = \{\boldsymbol{y} \in X \mid \boldsymbol{x} \leq \boldsymbol{y}\}$$

with a similar definition for $V_-(\boldsymbol{x})$. These cones are closed for the topology of pointwise convergence.

The following statement and was observed in Angenent (1988), and is related to the maximum principle for parabolic PDEs.

Theorem 14.2 (Strict Monotonicity of ζ^t) *For $\boldsymbol{x}, \boldsymbol{y} \in X$ with $\boldsymbol{x} < \boldsymbol{y}$ one has $\zeta^t(\boldsymbol{x}) \prec \zeta^t(\boldsymbol{y})$ for all $t > 0$.*

We will give a simple proof of this theorem in the appendix to this chapter ???. It is also a consequence of the Sturmian Lemma (see below), which was stated in Angenent (1988). Both ??? were communicated to the author by Sigurd Angenent. In Chapter AM, we defined the notion of crossing of two sequences $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^{\mathbb{Z}}$ in terms of their Aubry diagrams. We remind the reader that such *crossing* occurs when there is a $k \in \mathbb{Z}$ at which either $x_k - y_k$ and $x_{k+1} - y_{k+1}$ have opposite signs, or $x_k = y_k$ and $x_{k-1} - y_{k-1}$ and $x_{k+1} - y_{k+1}$ have opposite signs. We say that two sequences are *transverse* if they have no *tangency*, *i.e.* there is no $k \in \mathbb{Z}$ at which $x_k = y_k$ and $x_{k-1} - y_{k-1}$ and $x_{k+1} - y_{k+1}$ have same sign. We now define the *intersection index* $I(\boldsymbol{x}, \boldsymbol{y})$ to be the number of such crossings.

Lemma 14.3 (Sturmian Lemma) *Let $\boldsymbol{x}, \boldsymbol{y} \in X$ have different rotation numbers. If $\boldsymbol{x}, \boldsymbol{y}$ are not transverse, then for all sufficiently small $\varepsilon > 0$ $\varphi^{\pm\varepsilon}\boldsymbol{x}, \varphi^{\pm\varepsilon}\boldsymbol{y}$ are and:*

$$I\left(\varphi^{-\varepsilon}\boldsymbol{x}, \varphi^{-\varepsilon}\boldsymbol{y}\right) > I\left(\varphi^{\varepsilon}\boldsymbol{x}, \varphi^{\varepsilon}\boldsymbol{y}\right).$$

Otherwise, as long as $\boldsymbol{x}$ and $\boldsymbol{y}$ stay transverse, their intersection index does not change.

Figure of tangency???

Proof. See the appendix.

Corollary 14.4 *The sets CO, CO_ω, and X_{pq} are all invariant under the flow ζ^t, and so are their quotients by the action of $T = \tau_{0,1}$.*

Proof. The inequalities of the type $\boldsymbol{x} < \tau_{m,n}\boldsymbol{x}$, which define the sets CO and CO_ω are all preserved under ζ^t. The invariance of X_{pq} comes from the periodicity of the generating function S and its derivatives: when $\boldsymbol{x} \in X_{pq}$ the infinite dimensional vector field ∇W for the the ODE (14.1) is a sequence of period n (made of subsequences of length n equal to ∇W_{pq}).

15. The Gradient Flow and the Aubry-Mather Theorem

In this section, we show how the existence of CO orbits of all rotation numbers can be recovered from the monotonicity of the "gradient" flow ζ^t. From Lemma 9.2 and Corollary 14.4, we know that the set $\mathrm{CO}_\omega/\mathbb{Z}$ is compact and invariant under the flow ζ^t. Rest points of the flow in this set lift to CO orbits of rotation number ω. It turns out that, even though ζ^t is not the gradient flow of any function, we can still make it gradient like when restricted to the appropriate subsets.

Denote by $X^K = \{\boldsymbol{x} \in X \mid \sup_{k\in\mathbb{Z}} |x_k - x_{k-1}| < K\}$.

Theorem 15.1 *Let $C \subset X^K/\mathbb{Z}$ be a compact invariant set for the flow ζ^t. Then C must contain a rest point for the flow. In particular $\mathrm{CO}_\omega/\mathbb{Z}$ contains a restpoint and thus the map has a CO orbit of rotation number ω.*

Proof. Assume, by contradiction, that there are no rest points in C. We show that, for some large enough N, the truncated energy function $W_N = \sum_{-N}^{N} S(x_k, x_{k+1})$ is a strict Lyapunov function for the flow ζ^t on C. More precisely, we find a real $a > 0$ such that $\frac{d}{dt} W_N(\boldsymbol{x}) < -a$ for all $\boldsymbol{x}$ in C. This immediately yields a contradiction since on one hand W_N decreases to $-\infty$ on any orbit in C, on the other hand, the continuous W_N is bounded on the compact K. To show that W_N is a Lyapunov function for some N, we start with:

Lemma 15.2 *Let C be as in Theorem 15.1. Suppose that there are no rest points in C. Then, there exist a real $\varepsilon_0 > 0$, a positive integer N_0 such that, for all $\boldsymbol{x} \in C$*

$$N \geq N_0 \Rightarrow \forall j \in \mathbb{Z}, \quad \sum_{j}^{j+N} (\nabla W(\boldsymbol{x})_k)^2 > \varepsilon_0.$$

Proof. Suppose by contradiction that there exist sequences j_n, N_n and $\boldsymbol{x}^{(n)}$ with $N_n \to \infty$ such that

$$\sum_{j_n}^{j_n+N_n} \left(\nabla W(\boldsymbol{x}^{(n)})_k\right)^2 \to 0. \tag{15.1}$$

Let $m(n) = -j_n - [N_n/2]$ where $[\cdot]$ is the integer part function, and let $\boldsymbol{x}'^{(n)} = \sigma^{m(n)} \boldsymbol{x}^{(n)}$. This new sequence $\boldsymbol{x}'^{(n)}$ is still in C, and satisfies:

$$\sum_{k=-[N_n/2]}^{N_n-[N_n/2]} \left(\nabla W(\boldsymbol{x}^{(n)})_k\right)^2 \to 0 \quad \text{as } n \to \infty.$$

By compactness of C, it has a subsequence that converges pointwise (i.e. in the product topology) to some $\boldsymbol{x}^\infty$ in C. Clearly, $\nabla W(\boldsymbol{x}^\infty)_k = \lim_{n\to\infty} \nabla W(\boldsymbol{x}'^{(n)})_k = 0$ for all k and thus $\boldsymbol{x}^\infty$ is a rest point, a contradiction. □

We now show that W_N is a strict Lyapunov function on C. By chain rule:

$$\begin{aligned}
\frac{d}{dt} W_N(\boldsymbol{x}) &= -\left[\sum_{-N}^{N} \partial_1 S(x_k, x_{k+1}) \nabla W(\boldsymbol{x})_k + \partial_2 S(x_k, x_{k+1}) \nabla W(\boldsymbol{x})_{k+1}\right] \\
&= -\left[\sum_{-N}^{N} \partial_1 S(x_k, x_{k+1}) \nabla W(\boldsymbol{x})_k + \sum_{-N+1}^{N+1} \partial_2 S(x_{k-1}, x_k) \nabla W(\boldsymbol{x})_k\right] \\
&= -\,\partial_1 S(x_{-N}, x_{-N+1}) \nabla W(\boldsymbol{x})_{-N} - \partial_2 S(x_N, x_{N+1}) \nabla W(\boldsymbol{x})_{N+1} \\
&\quad - \sum_{-N+1}^{N} (\nabla W(\boldsymbol{x})_k)^2
\end{aligned} \tag{15.2}$$

For all $\boldsymbol{x}$ in X^K, we have $|x_k - x_{k-1}| < K$ and hence, by periodicity, $S(x_{k-1}, x_k)$, its partial derivatives and thus ∇W_k are all bounded on that set. In particular, we can find some M depending only on K such that

$$|-\partial_1 S(x_{-N}, x_{-N+1}) \nabla W(\boldsymbol{x})_{-N} - \partial_2 S(x_N, x_{N+1}) \nabla W(\boldsymbol{x})_{N+1}| < M$$

for all $\boldsymbol{x}$ in X^K and all integer k. We claim that for $N > (M+2)N_0/(2\varepsilon_0)$ (where N_0, ε_0 are as in Lemma 15.2), W_N is a Lyapunov function. Indeed, $N \geq (p+1)N_0$ where $p > M/\varepsilon_0$ and we can split the sum $\sum_{-N+1}^{N} (\nabla W(\boldsymbol{x})_k)^2$ into p sums of length greater than N_0. By Lemma 15.2, each of these sums must be greater than ε_0, and thus the total sum must be greater than $M + 2\varepsilon_0$, making the expression in (15.2) less than $-2\varepsilon_0$. □

Remark 15.3 As in Chapter AM, we can derive from Theorem 15.1 the existence of Aubry-Mather sets of all rotation numbers. This proof does not yield the fact that the orbits found are minimizers. This apparent weakness may be an asset in considering possible generalizations of this theorem to higher dimensions (see ???). This proof is a variation of the one given in Golé (1992 b). It was inspired by arguments found in Koch & al. (1994) , who prove an interesting generalization of the Aubry-Mather Theorem for functions on lattices of any dimensions.

16. Ghost Circles

Definition 16.1 A subset $\Gamma \subset \mathbb{R}^{\mathbb{Z}}$ is a Ghost Circle, hereafter GC, if it is

1. strictly ordered: $\boldsymbol{x}, \boldsymbol{y} \in \Gamma \Rightarrow \boldsymbol{x} \prec \boldsymbol{y}$ or $\boldsymbol{y} \prec \boldsymbol{x}$.
2. invariant under the $\mathbb{Z}^2$ action (by $\tau_{m,n}$), as well as under the flow ζ^t,
3. closed and connected.

We think of the GC's as the surviving traces *in the sequence space* $\mathbb{R}^{\mathbb{Z}}$ of the invariant circles of the twist map as one follows a one parameter family of maps away from a completely integrable map. We will see in the Section 17 that GC's can be constructed by bridging the gaps of the Aubry-Mather sets (identified to their corresponding subsets of rest points in $\mathbb{R}^{\mathbb{Z}}$) with connecting orbits of the gradient flow ζ^t.

Any sequence $\boldsymbol{x}$ in a GC Γ is CO: since $\tau_{m,n}\boldsymbol{x}$ must also lie in Γ, which is ordered, we must have $\boldsymbol{x} \prec \tau_{m,n}\boldsymbol{x}$ or $\tau_{m,n}\boldsymbol{x} \prec \boldsymbol{x}$. Moreover, the fact that Γ is ordered implies, by Lemma 13.2, that all sequences in Γ have same rotation number. We will call this $\rho(\Gamma)$, the *rotation number of the ghost circle.*

Proposition 16.3 *Let Γ be a ghost circle.*
a) The coordinate projection map $\mathbb{R}^{\mathbb{Z}} \mapsto \mathbb{R}$ defined by $\boldsymbol{x} \mapsto x_0$ induces a homeomorphism of Γ to $\mathbb{R}$. The corresponding projection map $\mathbb{R}^{\mathbb{Z}}/Z \mapsto \mathbb{R}/Z$ induces a homeomorphism between $\Gamma/\mathbb{Z}$ and the circle.
b) The set of ghost circles is closed in the Hausdorff topology of closed sets of $\mathbb{R}^{\mathbb{Z}}$, and it is compact in $\mathrm{CO}_{[a,b]}/\mathbb{Z}$. The rotation number on GCs is continuous in this topology.

We will see in GClemmonlimgcthat part b) of this proposition can be improved: monotonic (for the order on GCs defined at the end of this section) sequences of GCs with bounded rotation numbers converge uniquely.

Proof of Proposition 16.3 We show that, for any $\boldsymbol{x}, \boldsymbol{y}$ in Γ, the projection $\delta : \boldsymbol{x} \mapsto x_0$ defines a homeomorphism from $[\boldsymbol{x}, \boldsymbol{y}] \cap \Gamma$ to the interval $[x_0, y_0]$ in $\mathbb{R}$. As before, we give $\mathbb{R}^{\mathbb{Z}}$ the product topology. The projection map δ is continuous and the set $[\boldsymbol{x}, \boldsymbol{y}]$ is compact, by Tychonov Theorem, as a product of closed intervals. Clearly δ preserves the strict order: $\boldsymbol{x} \prec \boldsymbol{y} \Rightarrow x_0 < y_0$ and hence it is one to one on Γ. Take any two points $\boldsymbol{x} \prec \boldsymbol{y}$ in Γ. As a continuous injection, the map δ defines a homeomorphism on the compact set $\Gamma \cap [\boldsymbol{x}, \boldsymbol{y}]$. We show that $\delta(\Gamma \cap [\boldsymbol{x}, \boldsymbol{y}]) = [\delta(\boldsymbol{x}), \delta(\boldsymbol{y})]$. For this, it suffices to show that $\Gamma \cap [\boldsymbol{x}, \boldsymbol{y}]$ is connected. Suppose not and $\Gamma \cap [\boldsymbol{x}, \boldsymbol{y}] = A \cup B$ where A and B are closed and disjoint in $\Gamma \cap [\boldsymbol{x}, \boldsymbol{y}]$. There are two possibilities: either both $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to the same set, say A or else $\boldsymbol{x} \in A, \boldsymbol{y} \in B$. In the first case, we could write Γ as the union of two disjoint closed sets:

$$\Gamma = [(V_-(\boldsymbol{x}) \cap \Gamma) \cup A \cup (V_+(\boldsymbol{y}) \cap \Gamma)] \bigcup_{\neq} B,$$

a contradiction since Γ is connected. The other case yields the same contradiction. Since Γ is ordered, any bounded open ball for the product topology intersects Γ inside an interval $[\boldsymbol{x}, \boldsymbol{y}]$. Hence what we have shown above implies in particular that δ is a local homeomorphism on Γ. To show that it is a global homeomorphism, it suffices to show that it is onto. Since Γ is τ–invariant, if $\boldsymbol{x}$ is a point of Γ, then $\tau_{m,0}\boldsymbol{x}$ is as well, and hence the set $\{x_0 + m \mid m \in \mathbb{Z}\}$ is in $\delta(\Gamma)$. By what we proved above, all the points in between are also in $\delta(\Gamma)$ and hence δ is onto $\mathbb{R}$.

This proves a). To prove b), note that if $\Gamma_k \to \Gamma$ as $k \to \infty$ then any point $\boldsymbol{x} \in \Gamma$ is limit (in the product topology of $\mathbb{R}^{\mathbb{Z}}$) of points $\boldsymbol{x}^{(k)} \in \Gamma_k$. Since $\tau_{m,n}$ and the flow ζ^t are continuous, Γ must be invariant under these maps. "Close" and "connected" are adjectives that also behave well under Hausdorff limits. Finally, to see that Γ is strictly ordered, note that if $\boldsymbol{x}, \boldsymbol{y}$ are in Γ, we can find sequences $\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)} \in \Gamma_k$ with $\boldsymbol{x} = \lim \boldsymbol{x}^{(k)}, \boldsymbol{y} = \lim \boldsymbol{y}^{(k)}$. By restricting to a subsequence, we can assume $\boldsymbol{x}^{(k)} \prec \boldsymbol{y}^{(k)}$ for all k. Since Γ_k is strictly ordered and ζ^t-invariant, we must have $\zeta^{-t}\boldsymbol{x}^{(k)} < \zeta^{-t}\boldsymbol{y}^{(k)}$ and hence $\zeta^{-t}\boldsymbol{x} \leq \zeta^{-t}\boldsymbol{y}$. The strict monotonicity of the flow now implies: $\boldsymbol{x} \prec \boldsymbol{y}$. The continuity of the rotation number is a direct consequences of the continuity o on CO sequences, given by Lemma 9.1. □

It follows from this proposition that any GC has a parametrization $\xi \in \mathbb{R} \mapsto x(\xi) \in \Gamma$ of the form

$$x(\xi) = (\cdots, x_{-1}(\xi), \xi, x_1(\xi), x_2(\xi), \cdots). \tag{16.1}$$

where the $x_j(\xi)$ are strictly increasing and continuous functions of ξ. In particular $\xi \mapsto x_1(\xi)$ is a homeomorphism of $\mathbb{R}$. Invariance of Γ under the $\mathbb{Z}^2$ action τ implies that $x_j(\xi + 1) \equiv x_j(\xi) + 1$, so that the x_j define homeomorphisms of the circle as well; τ-invariance also implies that $x_2(\xi) = x_1(x_1(\xi))$, and more generally that the x_n are iterates of x_1.

Any GC projects naturally to a circle $\pi\Gamma$ in the annulus, where the projection $\pi : \mathbb{R}^{\mathbb{Z}} \to \mathbf{A}$ is defined by

$$\pi(\boldsymbol{x}) = (x_0, -\partial_1 S(x_0, x_1))$$

Proposition 16.3 *Let Γ be a GC for the twist map f. Then $\pi\Gamma$ and $f(\pi\Gamma)$ are periodic graphs over the $\boldsymbol{x}$ axis in $\mathbb{R}^2$. More precisely they are graphs of functions $\varphi(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ such that there is a constant $L < \infty$, depending only on the map, and, where the derivatives are defined,*

$$\varphi'(\boldsymbol{x}) \geq -L, \qquad \psi'(\boldsymbol{x}) \leq L.$$

Proof. If one parametrizes Γ as in (16.1) , then $\pi\Gamma$ is the graph of

(16.2) $$y = -\partial_1 S(\xi, x_1(\xi)) = \varphi(\xi).$$

The image $f(\pi\Gamma)$ is the graph of $y = \partial_2 S(x_{-1}(\xi), \xi)$. We now give a proof of the Lipschitz estimate. Using the parametrization of the projection of our GC as in (16.2) , it is enough to prove that the derivative of φ is bounded below. The same proof would hold for the estimate for the image $f(\pi\Gamma)$ of our circle. Applying the chain rule to (16.2) , we find:

$$\varphi' = -\partial_{11} S - \partial_{12} S \cdot \frac{dx_1}{d\xi} \geq -\partial_{11} S.$$

This last term is bounded below by our assumption on the second derivative of S. □

We end this section by giving a condition that insures that GCs do not intersect. We can define a partial ordering on GC's as follows. Let Γ_1, Γ_2 be GCs: then we say $\Gamma_1 \prec \Gamma_2$ if

(i) for all $\boldsymbol{x} \in \Gamma_1, \boldsymbol{x}' \in \Gamma_2$ one has $\boldsymbol{x} \pitchfork \boldsymbol{x}'$ and $I(\boldsymbol{x}, \boldsymbol{x}') = 1$;

(ii) $\rho(\Gamma_1) < \rho(\Gamma_2)$, i.e. $\rho(\boldsymbol{x}) < \rho(\boldsymbol{x}')$.

Lemma 16.4 Graph Ordering Lemma *If $\Gamma_1 \prec \Gamma_2$ then the circle $\pi\Gamma_1$ lies below $\pi\Gamma_2$.*

Proof. Let $x_n^{(j)}(\xi)$ be parametrizations of the form (16.1) for Γ_j ($j = 1, 2$). Then $\pi\Gamma_j$ is the graph of $\varphi_j(\xi) = (\xi, -\partial_1 S(\xi, x_1^{(j)}(\xi))$. We claim that $x_1^{(1)}(\xi) < x_1^{(2)}(\xi)$ for all ξ. Indeed, for a given ξ the sequences $x_n^{(1)}(\xi)$ and $x_n^{(2)}(\xi)$ intersect at site $n = 0$. Since they are transverse, we must have $x_1^{(1)}(\xi) \neq x_1^{(2)}(\xi)$; by comparing rotation numbers we then get $x_1^{(1)}(\xi) < x_1^{(2)}(\xi)$. By combining this inequality with the twist condition $\partial_{12} S < 0$ we then conclude that $\varphi_1(\xi) < \varphi_2(\xi)$, as claimed. □

Exercise 16.5 Prove that the set of $\boldsymbol{x}$ sequences corresponding to orbits of an nontrivial invariant circle for the map is a GC. (If the map has a transitive invariant circle of rotation number ω, then its associated GC is the only GC with rotation number ω (Golé (1992 a), Lemma 4.22). We conjecture that this remains true when the invariant circle is not transitive (i.e., of Denjoy type)).

17. Construction of Ghost Circles

This section will show that GCs are plentifull. In the first subsection we construct GCs whose projection passes through any given Aubry-Mather set. The next subsection will specialize to GCs with rational rotation numbers. For generical twist maps, we construct smooth GCs containing periodic minimizers. In Section 18 we will refine this construction to obtain ordered sets of GCs, whose projections do not intersect.

A. Ghost Circles Through Any Aubry-Mather Sets

Let M_ω the minimal, recurrent Aubry-Mather set of rotation number ω. It corresponds bijectively to the set, call it Σ_ω of $\boldsymbol{x}$ sequences of orbits in M_ω. By Aubry's Fundamental Lemma 10.2, Σ_ω is a completely ordered subset of CO_ω. If $\boldsymbol{x}$ is a recurrent minimizer, than so is $\tau_{m,n}\boldsymbol{x}$ for any $m, n \in \mathbb{Z}$, so Σ_ω is invariant under τ. Each point of Σ_ω corresponds to an orbit of F, and thus is a rest point of ζ^t. In Golé (1992 a), we proved the following theorem:

Theorem 17.1 *The set Σ_ω is included in a ghost circle Γ, and hence the Aubry-Mather set M_ω is included in the projection $\pi\Gamma$ of a ghost circle.*

Proof (Sketch) Σ_ω is a Cantor set whose complementary gaps are included in order intervals of the type $]\boldsymbol{x}, \boldsymbol{y}[$ where $\boldsymbol{x}, \boldsymbol{y} \in \Sigma_\omega$. A theorem of Dancer and Hess (1991) on monotone flows implies that, in conditions that are satisfied in the present case, if $\boldsymbol{x} \prec \boldsymbol{y}$ are two rest points for the strictly monotone flow ζ^t and there is no other restpoint in $[\boldsymbol{x}, \boldsymbol{y}]$ then there must be a *monotone orbit* ($i.e.$ completely ordered) of ζ^t joining $\boldsymbol{x}$ and $\boldsymbol{y}$. Hence we can bridge all the gaps of Σ_ω with ordered orbits of ζ^t, taking care to do so in an equivariant way with respect to the τ action. The resulting set is a GC. □

B. Smooth, Rational Ghost Circles

We now build rational Ghost Circles by piecing together the unstable manifolds of mountain pass points for $W_{p,q}$ in $X_{p,q}$. This construction will be crucial when we build disjoint GCs in Section 18. Let $\omega = p/q$ be given. Beginning here and throughout Sections 18 and 19 , we shall assume the following:

$$\textit{For any } p/q \in \mathbb{Q}\ W_{pq} \textit{ is a Morse–function on } X_{pq}. \tag{17.1}$$

This is a generic condition on twist maps, as will be proven in Proposition STMPpropgeneric. Since a GC consists of CO sequences we may assume that p and q have no common divisor (see Exercise AMexowellordering). Let $x \in X_{pq}$ be a critical point of W_{pq}. The second derivative of W_{pq} at $\boldsymbol{x}$ is a *Jacobi matrix*: it is tridiagonal (with positive "corner" elements as well) with positive subdiagonal terms:

$$-\nabla^2 W_{pq}(\boldsymbol{x}) = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \cdots & \beta_q \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & 0 \\ 0 & \beta_2 & \alpha_3 & \ddots & \vdots \\ & & \ddots & \ddots & \beta_{q-1} \\ \beta_q & 0 & \cdots & \beta_{q-1} & \alpha_q \end{bmatrix}, \tag{17.2}$$

where $\alpha_j = -\partial_{22} S(x_{j-1}, x_j) - \partial_{11} S(x_j, x_{j+1})$, and $\beta_j = -\partial_{12} S(x_{j-1}, x_j) > 0$. Due to the Perron–Fröbenius theorem, the largest eigenvalue λ_0 of $-\nabla^2 W_{pq}(\boldsymbol{x})$ is simple, and the eigenvector $\xi = (\xi_j)$ corresponding to λ_0 can be chosen to be positive: $\xi_j > 0, j = 1, \dots, q$. (See Angenent (1988), Proposition 3.2 and Lemma 3.4). If $\boldsymbol{x}$ is a critical point of index 1, there exist two orbits $\alpha_\pm(\boldsymbol{x}; t), t \in \mathbb{R}$ of the gradient flow ζ^t of W_{pq} with $\alpha_\pm(\boldsymbol{x}; t) \to x$ as $t \to -\infty$, and with

$$\alpha_\pm(\boldsymbol{x}; t) = \boldsymbol{x} \pm e^{\lambda_0 t} \xi + o\left(e^{\lambda_0 t}\right).$$

These two orbits, together with $\boldsymbol{x}$ itself, form the unstable manifold of $\boldsymbol{x}$. The orbits $\alpha_\pm(\boldsymbol{x}; t)$ are monotone, α_+ being increasing, and α_- decreasing; since $\tau_{\pm 1,0}\boldsymbol{x} = \boldsymbol{x} \pm 1$ are also critical points, we have $x - 1 \leq \alpha_\pm(\boldsymbol{x}; t) \leq x + 1$ so that the $\alpha_\pm(t)$ are bounded. Hence the limits

$$\omega_\pm(\boldsymbol{x}) = \lim_{t \to \infty} \alpha_\pm(\boldsymbol{x}; t)$$

exist; they are critical points of W_{pq}, and there are no other critical points $\boldsymbol{y}$ with $\omega_-(\boldsymbol{x}) < \boldsymbol{y} < \boldsymbol{x}$ or $\boldsymbol{x} < \boldsymbol{y} < \omega_+(\boldsymbol{x})$. If $\boldsymbol{y} > \boldsymbol{x}$ is another critical point, then $\boldsymbol{y} \geq \omega_+(\boldsymbol{x})$. refangenentperiodic, Theorem 1, shows that the points $\omega_\pm(\boldsymbol{x})$ are local minima of W_{pq}. We now show that the orbits $\alpha_\pm(\boldsymbol{x}; t)$ converge to these points along their "slow stable manifold", tangent to the largest eigenvalue of $-\nabla^2 W_{pq}(\omega_\pm(\boldsymbol{x}))$. Indeed, since $\omega_\pm(\boldsymbol{x})$ are minima, all the eigenvalues are negative, and thus the largest one has the smallest modulus. All orbits in the stable manifold of $\omega_\pm(\boldsymbol{x})$ except for a finite number that are tangent to the eigenspaces of the other eigenvalues, are tangent to this "slow stable manifold". But the other eigenvectors are in different orthants than the positive or negative ones (Angenent (1988)). Hence $\alpha_\pm(\boldsymbol{x};t)$, which are in the positive or negative orthant of $\omega_\pm(\boldsymbol{x})$, must converge to $\omega_\pm(\boldsymbol{x})$ tangentially to the eigenvector of largest eigenvalue.

To construct a GC in W_{pq} we first consider the set of critical points such a GC must contain.

Definition 17.2 A subset $\mathcal{A} \subset X_{pq}$ is a *skeleton* if the following hold.

S_1 $\mathcal{A}$ consists of critical points of W_{pq} with Morse index ≤ 1,

S_2 $\mathcal{A}$ is invariant under the $\mathbb{Z}^2$ action τ,

S_3 $\mathcal{A}$ is completely ordered.

A skeleton $\mathcal{A}$ is *maximal* if the only skeleton $\mathcal{A}'$ with $\mathcal{A} \subset \mathcal{A}' \subset X_{pq}$ is $\mathcal{A}$ itself.

Lemma 17.3 *A maximal skeleton $\mathcal{A}$ for W_{pq} exists.*

Proof. Choose r, s with $rp + qs = 1$ and define $\tau = \tau_{r,s}$ as in Exercise AMexowellordering. By Aubry's fundamental lemma the set $\mathcal{A}_0$ of absolute minimisers of W_{pq} is a skeleton. We fix some element $\boldsymbol{x} \in \mathcal{A}_0$. Any skeleton $\mathcal{A} \supset \mathcal{A}_0$ is completely determined by

$$\mathcal{B} = \mathcal{A} \cap [\boldsymbol{x}, \tau(\boldsymbol{x})] = \{\boldsymbol{z} \in \mathcal{A} : \boldsymbol{x} < \boldsymbol{z} < \tau(\boldsymbol{x})\}.$$

Indeed, given $\mathcal{B}$ we can reconstruct $\mathcal{A}$ as follows:

$$\mathcal{A} = \bigcup_{j=-\infty}^{\infty} \tau^j(\mathcal{B}). \tag{17.3}$$

Conversely, any ordered set of critical points $\mathcal{B} \subset [x, \tau(\boldsymbol{x})]$ determines a skeleton $\mathcal{A} \supset \mathcal{A}_0$ by ((17.3)). The closed order interval $[x, \tau(\boldsymbol{x})]$ is compact and W_{pq} is a Morse function, so there are only finitely many critical points in $[x, \tau(\boldsymbol{x})]$. We can therefore choose a maximal ordered set of critical points $\mathcal{B} \subset [x, \tau(\boldsymbol{x})]$ and be sure that the corresponding $\mathcal{A}$ is a maximal skeleton. □

Lemma 17.4 Mountain Pass Lemma *If the skeleton $\mathcal{A}$ is maximal, then every other point (according to the order) is a local minimum; the remaining points are minimaxes.*

Proof. If $\boldsymbol{x} < \boldsymbol{y}$ are consecutive elements of $\mathcal{A}$ then we must show that $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot both be local minima, while one of them must be a local minimum.

Step 1. If $\boldsymbol{x}$ and $\boldsymbol{y}$ both are local minima then the following standard minimax argument shows that there is a third critical point with index 1 between $\boldsymbol{x}$ and $\boldsymbol{y}$. Define $Q = [\boldsymbol{x}, \boldsymbol{y}]$ and consider

$$Q_\gamma = \{\boldsymbol{z} \in Q : W_{pq}(\boldsymbol{z}) \le \gamma\}$$

Each Q_γ is compact, and if $\gamma > \max W_{pq}\big|_Q$ then $Q_\gamma = Q$ is connected. On the other hand, Q_{γ_0} with $\gamma_0 = \max(W_{pq}(\boldsymbol{x}), W_{pq}(\boldsymbol{y}))$ is not connected, since $\boldsymbol{x}$ and $\boldsymbol{y}$ are local minima of W_{pq}. Consider

$$\gamma_1 = \inf\{\gamma > \gamma_0 : Q_\gamma \text{ connected}\}.$$

By compactness Q_{γ_1} itself is connected, and hence $\gamma_1 > \gamma_0$. Suppose there is no critical point of W_{pq} in $]\boldsymbol{x}, \boldsymbol{y}[$. Recall that Q is forward invariant under the gradient flow: $\zeta^t(Q) \subset Q$ for $t \geq 0$. By compactness of $Q_{\gamma_1} = \cap_{\gamma > \gamma_1} Q_\gamma$ there is an $\varepsilon > 0$ such that $\varphi^1(Q_{\gamma_1}) \subset Q_{\gamma_1 - \varepsilon}$, which implies that $Q_{\gamma_1-\varepsilon}$ is also connected, a contradiction. Hence there is at least one critical point $z \in]\boldsymbol{x}, \boldsymbol{y}[$, with $W_{pq}(\boldsymbol{z}) = \gamma_1$. If the Morse index of all such $\boldsymbol{z}$ were 2 or more, then the Morse Lemma TOPOlemmorsewould show that Q_γ with γ slightly less than γ_1 would still be connected — so the index of at least one such $\boldsymbol{z}$ is 1. But now we have a contradiction: if $\boldsymbol{x}$ and $\boldsymbol{y}$ are both local minima, then there is a minimax point $\boldsymbol{z} \in]\boldsymbol{x}, \boldsymbol{y}[$ and $\mathcal{A} \cup \{\tau_{m,n}\boldsymbol{z} : m, n \in \mathbb{Z}\}$ is a skeleton; this cannot be since $\mathcal{A}$ was maximal.

Step 2. Next we show that either $\boldsymbol{x}$ or $\boldsymbol{y}$ is a local minimum.If $\boldsymbol{x}$ is not a local minimum, then $\omega_+(\boldsymbol{x}) = \lim_{t\to\infty} \alpha_+(\boldsymbol{x}; t)$ is a local minimum. But $\omega_+(\boldsymbol{x}) \leq \boldsymbol{y}$, so $\omega_+(\boldsymbol{x}) = \boldsymbol{y}$, and we find that $\boldsymbol{y}$ must be a local minimum. Likewise, if $\boldsymbol{y}$ is not a local minimum, then $\boldsymbol{x} = \omega_-(\boldsymbol{y})$ must be one. □

We have all the ingredients necessary to show the following, which was proven in a slightly different form in Golé (1992 a), Theorem 3.6.

Theorem 17.5 *Assume W_{pq} is a Morse function. If $\mathcal{A}$ is a maximal skeleton, then*

$$\Gamma_{\mathcal{A}} = \{\alpha_\pm(\boldsymbol{x}; t) : t \in \mathbb{R}, \boldsymbol{x} \in \mathcal{A} \text{ is a minimax}\} \cup \mathcal{A}$$

is a C^1 Ghost Circle.

Proof. It is simple to check that, by maximality, $\Gamma_{\mathcal{A}}$ is connected, and a ghost circle. As a union of unstable manifolds, $\Gamma_{\mathcal{A}}$ is smooth except perhaps where different unstable manifold meet, at the minima. But we showed above how the orbits $\alpha_\pm(\boldsymbol{x}; t)$ must converge tangentially to the one dimensional eigenspace in the positive-negative cone of the minima. Hence the GC constructed is also smooth at the minima. □

Exercise 17.6 Check that $\Gamma_{\mathcal{A}}$ is indeed a GC.

18. Construction of disjoint Ghost Circles

We now arrive at the main result of this Chapter, which provides a vertical ordering of Aubry-Mather sets:

Theorem 18.1 (Ordering of Aubry-Mather Sets) *Given any interval $[a, b]$ in $\mathbb{R}$ there is a family of nontrivial circles $C_\omega, \omega \in [a, b]$ in the cylinder such that:*
(a) Each C_ω is the projection of a GC Γ_ω and hence is a graph over $\{y = 0\}$ (as is $f(C_\omega)$).
(b) The C_ω are mutually disjoint and if $\omega > \omega'$, C_ω is above $C_{\omega'}$.
(c) Each C_ω contains the Aubry-Mather set M_ω of recurrent minimizer of rotation number ω.

This section and the next two are devoted to the proof of this theorem. We will first construct, in this section and next one, finite families of rational ghost circles. In Section 20, we will take limits of such families and conclude the proof of the theorem.

Let $\omega_1, \ldots, \omega_k$ be distinct rational numbers. The construction of the preceding section provides us with maximal skeletons $\mathcal{A}_1, \ldots, \mathcal{A}_k$ and corresponding GC's $\Gamma_{\mathcal{A}_1}, \ldots, \Gamma_{\mathcal{A}_k}$. It is not immediatly clear from this construction that the projections $C_j = \pi\Gamma_{\mathcal{A}_j}$ are disjoint. In this section we show that the skeletons can be chosen so that the C_j are indeed disjoint.

Definition 18.2 A family of skeletons $\mathcal{A}_j \subset X_{p_j q_j}$ is minimally linked if any pair $x \in \mathcal{A}_i, y \in \mathcal{A}_j$ with $i \neq j$ is transverse with $I(x, y) = 1$.

Theorem 18.3 Disjointness Theorem *If $\mathcal{A}_j \subset X_{p_j q_j}$ is a minimally linked family of maximal skeletons, then the projected GC's $C_j = \pi\Gamma_{\mathcal{A}_j}$ are disjoint.*

Proof. Order the $\mathcal{A}_j$ so that their rotation numbers $\rho_j = p_j/q_j$ are increasing. Then we claim that

$$\Gamma_{\mathcal{A}_1} \prec \Gamma_{\mathcal{A}_2} \prec \Gamma_{\mathcal{A}_3} \prec \cdots \prec \Gamma_{\mathcal{A}_k}. \tag{18.1}$$

Disjointness of the projected GCs then follows directly from the Graph Ordering Lemma. To see why (18.1) holds, we consider any pair $\boldsymbol{x}^{(i)} \in \Gamma_{\mathcal{A}_i}, \boldsymbol{x}^{(j)} \in \Gamma_{\mathcal{A}_j}$ and assume that they are not transverse. By the Sturmian Lemma

$$I\left(\zeta^t \boldsymbol{x}^{(i)}, \zeta^t \boldsymbol{x}^{(j)}\right) > 1 \tag{18.2}$$

for all those $t < 0$ at which $\zeta^t \boldsymbol{x}^{(i)} \pitchfork \zeta^t \boldsymbol{x}^{(j)}$. But $\lim_{t\to-\infty} \zeta^t \boldsymbol{x}^{(l)} = \boldsymbol{y}^{(l)}$ for some $\boldsymbol{y}^{(l)} \in \mathcal{A}_l$ $(l = i, j)$. Since the $\mathcal{A}_l$ are minimally linked we must have $I(\boldsymbol{y}^{(i)}, \boldsymbol{y}^{(j)}) = 1$, which contradicts (18.2) . □

Theorem 18.4 *For any k-tuple $\omega_1, \ldots, \omega_k$ of rational numbers there exists a minimally linked family of skeletons $\mathcal{A}_1, \ldots, \mathcal{A}_k$ such that each $\mathcal{A}_j$ is a maximal skeleton.*

This theorem, combined with the Disjointness Theorem, provides us with a disjoint family of circles $C_j = \pi\Gamma_{\mathcal{A}_j}$ in the annulus. The construction of the $\mathcal{A}_j$'s will be such that they automatically contain the absolute minimizers of W_{ω_i}, which by Proposition 10.4 are the minimal energy orbits of Aubry–Mather. In our proof of Theorem 18.4 we begin with constructing a maximal k-tuple of skeletons, and then show that each skeleton in this k-tuple is maximal.

Proof of Theorem 18.4 Let $\mathcal{M}_j$ be the set of absolute minimizers of W_{ω_j} on $X_{p_j q_j}$. Aubry's fundamental lemma implies that $\mathcal{M}_1, \ldots, \mathcal{M}_k$ is a minimally linked family of skeletons. As in the proof of Lemma 17.3one easily finds a maximal k-tuple of skeletons $\mathcal{A}_1, \ldots, \mathcal{A}_k$ with $\mathcal{M}_j \subset \mathcal{A}_j$, by observing that there are only finitely many of such extensions. We shall now show that each $\mathcal{A}_j$ is a maximal skeleton.

Assume that one of the $\mathcal{A}_j$, say $\mathcal{A}_1$ is not maximal. Then there is a critical point $z \in W_{p_1 q_1}$ with index 0 or 1, such that $\mathcal{A}_1 \cup \{z\}$ is completely ordered. In particular, there must exist a couple of adjacent critical points $\boldsymbol{x} < \boldsymbol{y}$ in $\mathcal{A}_1$ with $\boldsymbol{z} \in]\boldsymbol{x}, \boldsymbol{y}[$. We must deal with two different cases:

A. Both $\boldsymbol{x}$ and $\boldsymbol{y}$ are local minima of $W_{p_1 q_1}$.

B. At least one of the critical points $\boldsymbol{x}$ or $\boldsymbol{y}$ has index 1.

Case A. By a minimax argument we will show that there is a critical point between $\boldsymbol{x}$ and $\boldsymbol{y}$ which allows us to extend $\mathcal{A}_1$ to a larger skeleton $\mathcal{A}_1'$ for which $(\mathcal{A}_1', \ldots, \mathcal{A}_k)$ is still minimally linked. This would then contradict maximality of $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$, and thereby show that Case A cannot occur. To carry out the minimax argument we consider

$$\Omega = \{\boldsymbol{w} \in W_{p_1 q_1} : \boldsymbol{x} < \boldsymbol{w} < \boldsymbol{y}, \forall_{j \geq 2} \forall_{v \in \mathcal{A}_j} \boldsymbol{v} \pitchfork \boldsymbol{w} \text{ and } I(\boldsymbol{v}, \boldsymbol{w}) = 1\}.$$

and its closure $Q = \bar{\Omega}$. The Sturmian theorem again implies that Ω and hence Q is forward invariant under the gradient flow ζ^t. Also, as in Lemma ???, Q is compact, as are the sublevel sets $Q_\gamma = \{\boldsymbol{w} \in Q : W_{p_1 q_1}(\boldsymbol{w}) \leq \gamma\}$. To obtain a critical point other than $\boldsymbol{x}$ and $\boldsymbol{y}$ in Q we must show that not all the Q_γ's have the same topology. If $\gamma_0 = \max(W_{p_1 q_1}(\boldsymbol{x}), W_{p_1 q_1}(\boldsymbol{y}))$, then Q_{γ_0} is again not connected, since $\boldsymbol{x}$ and $\boldsymbol{y}$ are local minima. On the other hand we have

Lemma 18.5 *$Q = \bar{\Omega}$ is connected.*

Postponing the proof of this statement to the next section, we can now easily complete the minimax argument. Indeed, as in Lemma 17.4,

$$\gamma_1 = \inf(\gamma > \gamma_0 : Q_\gamma \text{ connected})$$

is a critical value of $W_{p_1 q_1}$, so there must be a third critical point $\boldsymbol{w} \in Q$. By the Sturmian Lemma $\boldsymbol{w}$ must lie in Ω, and it follows from the Morse lemma that $\boldsymbol{w}$ has index 1. Put

$$\mathcal{A}_1' = \mathcal{A}_1 \cup \{\tau_{m,n} \boldsymbol{w} : m, n \in \mathbb{Z}\}; \tag{18.3}$$

then $(\mathcal{A}_1', \ldots, \mathcal{A}_k)$ is a minimally linked family of skeletons extending $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$, and we have our contradiction.

Case B. Assume that $\boldsymbol{x}$ is not a local minimum, and put $\boldsymbol{w} = \omega_+(\boldsymbol{x})$. Then $\boldsymbol{w}$ is a critical point of $W_{p_1 q_1}$ and is therefore transverse to any $v \in \mathcal{A}_j$ with $j \geq 2$, by Lemma 3.5. We claim that $I(\boldsymbol{w}, \boldsymbol{v}) = 1$. Indeed, for $t \to -\infty$ we have $\alpha_+(\boldsymbol{x}; t) \to x$. Since $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$ is minimally linked, we find that for all t sufficiently large negative $\alpha_+(\boldsymbol{x}; t)$ and v are transverse with $I(\alpha_+(\boldsymbol{x}; t), v) = 1$. By the Sturmian Lemma $I(\alpha_+(\boldsymbol{x}; t), v)$ cannot increase, and since $\alpha_+(\boldsymbol{x}; t)$ and v have different rotation numbers $I(\alpha_+(\boldsymbol{x}; t), v) \geq 1$ for all t: hence $I(\alpha_+(\boldsymbol{x}; t), v) = 1$ for all t. Letting $t \to +\infty$ we get $I(\boldsymbol{w}, \boldsymbol{v}) = 1$, as claimed. Defining $\mathcal{A}_1'$ as in ((18.3)) we again get a larger minimally linked family of skeletons, a contradiction. If $\boldsymbol{x}$ is a local minimum then $\boldsymbol{y}$ cannot be one, and consideration of $\omega_-(\boldsymbol{y})$ leads to a similar contradiction. □

19. Proof of Lemma 18.5

We must show that $Q = \bar{\Omega}$ is connected. We shall do this by showing that any $\boldsymbol{w} \in \Omega$ can be connected to $\boldsymbol{x}$ via a path $\gamma : [0, 1] \to \Omega \cup \{\boldsymbol{x}\}$.

For any $j \in \mathbb{Z}$ and any $\boldsymbol{x}, \boldsymbol{w} \in X_{p_1 q_1}$ we put

$$A_j(\boldsymbol{x} : \boldsymbol{w}) = \{v_j : \boldsymbol{v} \in \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k\} \cap [x_j, w_j).$$

For simplicity we shall write $x \pitchfork \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$ when we mean that $x \pitchfork v$ for every $v \in \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$.

Proposition 19.1

(i) $A_j(\boldsymbol{x} : \boldsymbol{w})$ *is finite, for each* j.

(ii) $A_{j+q_1}(\boldsymbol{x} : \boldsymbol{w}) = A_j + p_1$.

(iii) If $\boldsymbol{z} \in X_{p_1 q_1}$ *and* $\boldsymbol{x} \le \boldsymbol{z} \le \boldsymbol{w}$, *then* $\boldsymbol{z} \pitchfork \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$, *if and only if they are tranverse in the index range* $0 \le j \le q_1$.

Proof. (i) : $W_{p_j q_j}$ is a Morse function. (ii) holds because $\boldsymbol{x}, \boldsymbol{w} \in X_{p_1 q_1}$ and the $\mathcal{A}_l$ are invariant under the action of $\tau_{m,n}$, $m, n \in \mathbb{Z}$. (iii) is a consequence of (ii). □

We define the *height* of $\boldsymbol{w}$ over $\boldsymbol{x}$ by

$$h(\boldsymbol{x} : \boldsymbol{w}) = \sum_{j=0}^{q_1-1} \#\left(A_j(\boldsymbol{x} : \boldsymbol{w})\right).$$

If the height $h(\boldsymbol{x} : \boldsymbol{w})$ vanishes then all the $A_j(\boldsymbol{x} : \boldsymbol{w})$ are empty and we can define $\gamma(t) = t\boldsymbol{w} + (1-t)\boldsymbol{x}$. Since $x_j \le \gamma_j(t) \le w_j$ for all j and $0 \le t \le 1$, it follows from part (iii) of our last proposition that $\gamma(t) \pitchfork \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$ for $0 \le t \le 1$, so that $\gamma(t)$ stays within Q. Call this a move of type 0.

We will now assume that $h(\boldsymbol{x} : \boldsymbol{w}) > 0$, and will show how to decrease it to zero. Suppose that for some l one has $w_l = v_l > x_l$ for some $\boldsymbol{v} \in \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$. Then there is an ε such that $0 < \varepsilon < w_l - x_l$ and $(w_l - \varepsilon, w_l) \cap A_l(\boldsymbol{x} : \boldsymbol{w})$ is empty and we can define

$$w'_j = \begin{cases} w_j - \varepsilon & \text{if } j = l \bmod q_1, \\ w_l & \text{otherwise.} \end{cases}$$

As before one can connect $\boldsymbol{w}$ and $\boldsymbol{w}'$ by $\gamma(t) = t\boldsymbol{w} + (1-t)\boldsymbol{w}'$ without leaving Q. Call this a move of type 1.

Assuming now that $w_i \ne v_i$ for all i, we will move the sequence $\boldsymbol{w}$ down by interpolating it linearly to:

$$z_i^{(l)} = \begin{cases} \max A_i(\boldsymbol{x} : \boldsymbol{w}) & \text{if } i = l \bmod q_1, \\ w_i & \text{otherwise} \end{cases}$$

for some judiciously chosen l. Call this is a move of type 2. Clearly $z^{(l)} \in X_{p_1 q_1}$ and $x \le z^{(l)} \le w$, $z^{(l)} = z^{(l+q)}$ and $h(x : z_i^{(l)}) = h(\boldsymbol{x} : \boldsymbol{w}) - 1$. We need to show that for at least one $l \in \mathbb{Z}$, this move does not change the intersection index of $\boldsymbol{w}$ with the elements of $\mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$. Consider the set of elements in $\mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$ that are immediately below $\boldsymbol{w}$:

$$a_i^{s_i} =_{\text{def}} \max A_i(\boldsymbol{x} : \boldsymbol{w}).$$

Assume that, among the sequences a^{s_i} at least one has rotation number greater than that of $\boldsymbol{x}$ and pick the one, say a^{s_j} which has the largest rotation number (If all a^{s_i} have lower rotation number than $\boldsymbol{x}$, pick the one that has the lowest and proceed similarly). In the following, we only worry about the possible changes of intersection index in the range $0 \le j \le \boldsymbol{q}$. The periodicity condition (ii) of Proposition 19.1 insures that if there are changes of index, they must occur periodically. There are three cases (see Figure 19. 1) to consider:

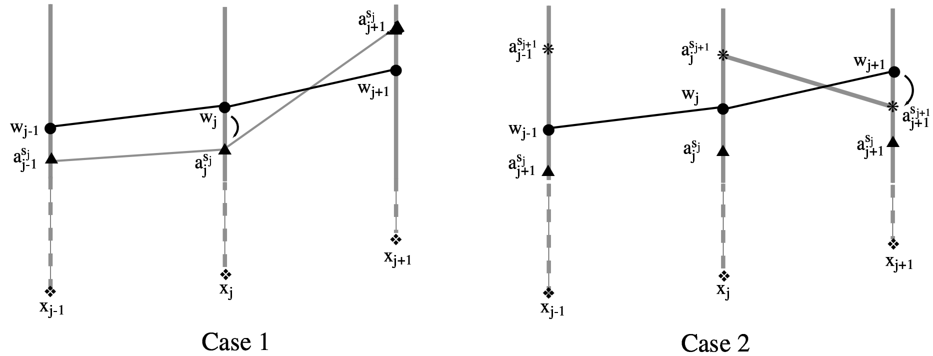


Fig. 19. 1. The two possible moves of type 2

Case 1: $a^{s_j}_{j+1} > w_{j+1}$
Choose $l = j$ and move $\boldsymbol{w}$ to $z^{(l)}$ as defined above. This could only change the intersection index of $\boldsymbol{w}$ with a^{s_j}. But in this case this intersection index remains the same: since $\rho(a^{s_j}) > \rho(w) = \rho(\boldsymbol{x})$, and $I(a^{s_j}, w) = 1$, we must have $a^{s_j}_{j-1} \leq a^{s_{j-1}}_{j-1} < w_{j-1}$. Hence the one crossing of $\boldsymbol{w}$ and a^{s_j}, which occured between j and $j+1$ is now moved to a crossing that occurs at j, with no other crossing introduced with this or any other sequence of $\mathcal{A}_2 \cup \cdots \cup \mathcal{A}_k$.
Case 2: $a^{s_j}_{j+1} < a^{s_{j+1}}_{j+1}$
Since by assumption $\rho(a^{s_{j+1}}) \leq \rho(a^{s_j})$, we must have $a^{s_{j+1}}_j > a^{s_j}_j$ and thus $a^{s_{j+1}}_j > w_j$, by maximality of $a^{s_j}_j$. Now choose $l = j+1$ and move $\boldsymbol{w}$ to $z^{(l)}$: the one crossing of $\boldsymbol{w}$ and $a^{s_{j+1}}$, which occured between j and $j+1$ is now moved to a crossing that occurs at $j+1$.
Case 3: $a^{s_j}_{j+1} = a^{s_{j+1}}_{j+1}$
The equality $a^{s_j}_i = a^{s_i}_i$ cannot be true for all $i > j$ since otherwise $\boldsymbol{w}$ and a^{s_j} would have same rotation number. Hence for some $i > j$, Case 1 or 2 must occur. Apply the procedure for these cases there.

Concatenating moves alternating between type 1 and 2, we get a curve in Q between $\boldsymbol{w}$ and and a sequence which has zero height. Concatenate this with a move of type 0 to get a curve in Q between $\boldsymbol{w}$ and $\boldsymbol{x}$. □

20. Proof of Theorem 18.1

Let $\omega_1, \omega_2, \cdots$ be an enumeration of the rational numbers in the interval (a, b).

Proposition 20.1 *There is a family of GCs* $\{\Gamma^{(n)}_1, \ldots, \Gamma^{(n)}_n\}$, *where* $\Gamma^{(n)}_j$ *has rotation number* ω_j, *and where* $\Gamma^{(n)}_i \prec \Gamma^{(n)}_j$ *if* $\omega_i < \omega_j$. *Each* $\Gamma^{(n)}_i$ *contains at least one minimizing periodic orbit of rotation number* ω_i, *and generically all of them.*

Proof. If one assumes that the map f is such that the Morse property 17.1 holds, then, according to Theorem 18.4, one can find a minimally linked family of maximal skeletons $\{\mathcal{A}^{(n)}_1, \ldots, \mathcal{A}^{(n)}_n\}$ such that $\mathcal{A}^{(n)}_j$ has rotation number ω_j and contains all the absolute minimizers of that rotation number. The corresponding GCs $\Gamma^{(n)}_i = \Gamma_{\mathcal{A}^{(n)}_i}$ then satisfy the required conditions.

In general, when the Morse property 17.1 is not satisfied, one can approximate f by smooth twist maps f_ε which do satisfy 17.1 (since this condition is generic); One thus obtains GCs $\Gamma_{j,\varepsilon}^{(n)}$, and by the compactness of the set of GCs with a fixed rotation number (Proposition 16.3) one can extract a convergent subsequence whose limit will then be a family $\{\Gamma_1^{(n)}, \ldots, \Gamma_n^{(n)}\}$ of GCs. But we need to make sure that limits of strictly ordered rational GCs stay strictly ordered. To see this, notice that the set $\Gamma_{i,\varepsilon}^{(n)} \times \Gamma_{j,\varepsilon}^{(n)}$ is, when $i \neq j$, included in:

$$\Omega_{ij} = \left\{(\boldsymbol{v}, \boldsymbol{w}) \in \mathrm{PCO}_{\omega_i} \times \mathrm{PCO}_{\omega_j} \ : \ \boldsymbol{v} \pitchfork \boldsymbol{w} \quad \text{and} \quad I(\boldsymbol{v}, \boldsymbol{w}) = 1\right\}$$

where PCO_ω is the set of periodic CO sequences of rotation number ω:

$$\mathrm{PCO}_{p/q} = CO_{p/q} \cap X_{pq}.$$

The set Ω_{ij} is, by the Sturmian lemma, positively invariant under the product gradient flow $\phi^t \times \phi^t$ corresponding to any twist map. In fact: $(\phi^t \times \phi^t)(Clos\ \Omega_{ij}) \subset (Int\ \Omega_{ij})$, as can easily be checked (i.e. $Clos\ \Omega_{ij}$ is an "attractor block" in the sense of Conley.) As Hausdorff limit of compact sets in Ω_{ij}, the set $\Gamma_i^{(n)} \times \Gamma_j^{(n)}$ is in $Clos\ \Omega_{ij}$. But, since it is both positively and negatively invariant under $\phi^t \times \phi^t$, $\Gamma_i^{(n)} \times \Gamma_j^{(n)}$ must in fact be in $Int\ \Omega_{ij}$ where the intersection number is well defined and always equal to 1. In other words, we have shown that, whenever $\omega_i < \omega_j$ one must have $\Gamma_i^{(n)} \prec \Gamma_j^{(n)}$. Finally, the set $\Gamma_i^{(n)}$ contains at least a minimizing periodic orbit, since the sets $\Gamma_{i,\varepsilon}^{(n)}$ contain by construction *all* the minimizing periodic orbits of period ω_i for f_ε, and limits of minimizers are minimizers. □

Rational C_ω's

We now construct the C_ω's of Theorem 18.1, starting with all the rational $\omega \in [a, b]$. Again, we use the compactness of the set of GCs: For each n the proposition provides us with GCs $\Gamma_1^{(n)}, \ldots, \Gamma_n^{(n)}$ with rotation numbers $\omega_1, \ldots, \omega_n$. By compactness we can extract a subsequence $\{n_j\}$ for which the $\Gamma_1^{(n_j)}$ converge as $j \to \infty$ to a GC of rotation number ω_1. Using compactness again, we can extract a further subsequence n'_j for which $\Gamma_1^{(n'_j)}$ and $\Gamma_2^{(n'_j)}$ both converge; repetition of this argument and application of the diagonal trick then finally gives a subsequence n''_j for which all $\Gamma_k^{(n''_j)}$ converge to some limiting GC $\Gamma_k^{(\infty)}$ (of rotation number ω_k) as $j \to \infty$. By the same argument as in the previous proposition, the limits $\Gamma_k^{(\infty)}$ satisfy $\Gamma_i^{(\infty)} \prec \Gamma_j^{(\infty)}$ whenever $\omega_i < \omega_j$. We then define $C_{\omega_k} = \pi \Gamma_k^{(\infty)}$ and by the Graph Ordering Lemma 16.4, the C_{ω_k}'s are disjoint. In the generic case, each $\Gamma_i^{(n)}$ contains all the periodic minimizers of rotation number ω_i, and hence so must the limit $\Gamma_i^{(\infty)}$. In the non generic case, $\Gamma_i^{(\infty)}$ must contain at least one periodic minimizer of the energy.

Irrational C_ω's

To complete our family of rational GCs with irrational ones, we once again take a limit. We could proceed in a way similar to what we did in order to get all rational GCs, but we would have to appeal to the axiom of choice (no diagonal tricks on uncountable sets!). To avoid this, we first prove a proposition of monotone convergence of GCs. We shall write $\Gamma_1 \preceq \Gamma_2$ if either $\Gamma_1 \prec \Gamma_2$ or $\rho(\Gamma_1) = \rho(\Gamma_2)$ and $\pi\Gamma_1$ is (not necessarily strictly) below $\pi\Gamma_2$.This last condition is equivalent to $x_1^{(1)}(\xi) \leq x_1^{(2)}(\xi)$ in the notation of the proof of the Graph Ordering Lemma 16.4.

Proposition 20.2 Monotone Convergence for Ghost Circles *Let $\Gamma^{(j)}$ be an increasing sequence of GCs, i.e. assume that*

$$\Gamma^{(1)} \preceq \Gamma^{(2)} \preceq \Gamma^{(3)} \preceq \cdots.$$

Assume also that the rotation numbers $\rho_j = \rho(\Gamma^{(j)})$ are bounded from above. Then there is a unique GC $\Gamma^{(\infty)}$ such that $\Gamma^{(j)} \to \Gamma^{(\infty)}$ as $j \to \infty$. Moreover, if $x^{(j)}(\xi)$ is the parametrization of $\Gamma^{(j)}$ with $x_0^{(j)}(\xi) \equiv \xi$, then the $x_k^{(j)}(\xi)$ converge monotonically and uniformly to $x_k^{(\infty)}(\xi)$, where $x^{(\infty)}(\xi)$ is the parametrization of $\Gamma^{(\infty)}$ with $x_0^{(\infty)}(\xi) \equiv \xi$

Of course, the corresponding theorem for decreasing sequences also holds. We postpone the proof of this proposition till the end of this section.

Assume now that we have constructed the rational GCs $\Gamma_k^{(\infty)}$ as above. For any number $\omega \in (a, b)$, rational or otherwise, we can then define two GCs $\Gamma_\omega^{\pm}$ as follows. Choose a sequence of rational numbers ω_{n_j} which increases monotonically to ω. The Monotone Convergence Theorem tells us that the limit of the corresponding GCs $\Gamma_{n_j}^{(\infty)}$ must exist. We denote this limit by Γ_ω^-. This procedure might produce an ambiguous definition of Γ_ω^-, since the result could depend on the choice of the sequence n_j: If one has two such sequences, n_j and n'_j, then the $\Gamma_{n_j}^{(\infty)}$ and $\Gamma_{n'_j}^{(\infty)}$ might have two different limits Γ and Γ'. However, one can take the union of the two sequences to obtain a third sequence n''_j, i.e. $\{n''_j\} = \{n_j\} \cup \{n'_j\}$. The $\omega_{n''_j}$ then also increase to ω, so that the $\Gamma_{n''_j}^{(\infty)}$ also must converge to some GC Γ''. Since n_j and n'_j are subsequences of n''_j, both sequences n_j and n'_j must produce the same limiting GC: hence $\Gamma = \Gamma' = \Gamma''$, and the definition of Γ_ω^- is independent of the choice of the n_j. We choose to define $C_\omega = \pi\Gamma_\omega^-$ (or $\pi\Gamma_\omega^+$, but with the same choice of $+$ or $-$ for all ω in order to avoid using the axiom of choice...).

We now check that, for ω irrational, the unique Aubry-Mather set M_ω of recurrent minimizers (see Proposition 12.9) is included in C_ω. We can take a sequence of periodic Aubry minimizing sequences $\boldsymbol{x}^k \in \Gamma_k^{(\infty)}$ where $\omega_k \nearrow \omega$ ($\searrow$ if one chose $C_\omega = \pi\Gamma^+$). Then $\boldsymbol{x}^k \to \boldsymbol{x}$, an Aubry minimizing sequence in Γ_ω^-. The orbit that $\boldsymbol{x}$ corresponds to is recurrent and minimizing, as limit of recurrent and minimizing orbits. Its closure, which is also included in C_ω, must be M_ω . From our definition of $\Gamma_\omega^{\pm}$, it is clear that:

$$\omega_i < \omega < \omega_j \Rightarrow \Gamma_i^{(\infty)} \prec \Gamma_\omega^- \preceq \Gamma_\omega^+ \prec \Gamma_j^{(\infty)},$$

for rational ω_i, ω_j and irrational ω. Hence the set formed by the rational GCs $\Gamma_k^{(\infty)}$ and the irrational ones Γ_ω is completely ordered according to their rotation numbers. By the Graph Ordering Lemma 16.4, the C_ω's (irrational and rational) that we have constructed are mutually disjoint. □

Remark 20.2 If ω is a rational number, Γ_ω^- is no longer necessarily in PCO_ω but is certainly in CO_ω. It may contain the sequences corresponding to homo(hetero)clinic orbits joining hyperbolic periodic orbits of rotation number ω. Hence we may (and, probably, generically do) have three distinct Ghost Circles $\Gamma_\omega^- \preceq \Gamma_\omega \preceq \Gamma_\omega^+$ for each rational ω where Γ_ω is $\Gamma_k^{(\infty)}$ for some k. We will call their projections C_ω^-, C_ω and C_ω^+ respectively. Instead of the set $\{C_\omega\}_{\omega\in[a,b]}$ of strictly non mutually intersecting curves that we have found in Theorem 18.1, one might prefer to consider the bigger set $\{C_\omega \cup C_\omega^+ \cup C_\omega^-\}_{\omega\in[a,b]}$. It is not hard to check that this is a closed set of GCs.

Proof of Proposition 20.2

Proof. It follows from the Graph Ordering Lemma 16.4 tthat the $x_k^{(j)}(\xi)$ are monotone in j. We have assumed that the rotation numbers of the $\Gamma^{(j)}$ are bounded, so we may as well assume they are bounded by some integer M. This bound implies for $l > 0$ that $x_l^{(j)}(\xi) \leq \xi + l(M+1)$, and in view of the monotonicity of the $x_l^{(j)}(\xi)$ they converge to some $x_l^{(\infty)}(\xi)$. For negative l one finds that $x_l^{(j)}(\xi) \geq \xi + l(M+1)$, so that the $x_l^{(j)}(\xi)$ decrease to some $x_l^{(\infty)}(\xi)$. Clearly $x_1^{(\infty)}(\xi)$ is a nondecreasing function of ξ. We shall show that it is strictly increasing, and continuous.

$x_1^{(\infty)}(\xi)$ *is strictly increasing.* Let $\xi < \eta$ be given. Then $t \mapsto \zeta^t(x^{(j)}(\xi))$ and $t \mapsto \zeta^t(x^{(j)}(\eta))$ both are on the GC $\Gamma^{(j)}$, so that they must be ordered in the same way for all $t \in \mathbb{R}$. At $t = 0$ we have

$$\xi = \zeta^t(x^{(j)}(\xi))_0 < \zeta^t(x^{(j)}(\eta))_0 = \eta$$

so this ordering must hold for all t. Upon taking the limit $j \to \infty$ we find that $\zeta^t(x^{(\infty)}(\xi)) \leq \zeta^t(x^{(\infty)}(\eta))$ holds for all t. By the Strict Monotonicity lemma we must have strict inequality for all t, unless we have equality for all t. Equality cannot happen of course, since $x_0^{(\infty)}(\xi) = \xi < \eta = x_0^{(\infty)}(\eta)$. Hence we have $x^{(\infty)}(\xi) < x^{(\infty)}(\eta)$; in particular $x_1^{(\infty)}(\xi) < x_1^{(\infty)}(\eta)$.

$x_1^{(\infty)}(\xi)$ *is continuous.* Since the $x_1^{(j)}(\xi)$ are monotonically increasing in both j and ξ, their limit is increasing and lower semicontinuous in ξ. Thus we only have to show that $x_1^{(\infty)}(\xi) = x_1^{(\infty)}(\xi + 0)$. Assume that $x_1^{(\infty)}(\xi) < x_1^{(\infty)}(\xi+0)$ and define $a = \{x_1^{(\infty)}(\xi) + x_1^{(\infty)}(\xi+0)\}/2$. Then there is a sequence $\delta_j \downarrow 0$ such that $x^{(j)}(\xi + \delta_j) = a$. As before we consider $\zeta^t\left(x^{(j)}(\xi+\delta_j)\right)$ and $\zeta^t\left(x^{(j)}(\xi)\right)$, and take the limit $j \to \infty$. Then, after passing to a subsequence if necessary, $\zeta^t\left(x^{(j)}(\xi+\delta_j)\right) \to \zeta^t(x^*)$ for some x^* with $x_0^* = \xi$ and $x_1^* = a$, while $\zeta^t\left(x^{(j)}(\xi)\right) \to \zeta^t\left(x^{(\infty)}(\xi)\right)$. Moreover we will have $\zeta^t(x^*) \geq \zeta^t\left(x^{(\infty)}(\xi)\right)$ for all t, again with either strict inequality for all t or equality for all t. But this contradicts the fact that at $t = 0$ we have $x_0^* = \xi = x_0^{(\infty)}(\xi)$ and $x_1^* = a > x^{(\infty)}(\xi)$. Thus $x_1^{(\infty)}(\xi)$ is indeed continuous. Since the $x_1^{(j)}(\xi)$ increase monotonically to $x_1^{(\infty)}(\xi)$, and since $x_1^{(\infty)}(\xi)$ is continuous, the convergence must be uniform (Dini's theorem). Therefore the $x_l^{(j)}(\xi)$, being iterates of $x_1^{(j)}(\xi)$ also converge uniformly.

One now easily verifies that $\Gamma^{(\infty)} = \{x^{(\infty)} : \xi \in \mathbb{R}\}$ is a GC, and that it is the limit in the Hausdorff metric of the $\Gamma^{(j)}$s. □

Exercise 20.3 Complete the sketch of this alternating conclusion of the proof of Theorem 18.1: For each $\rho = (\omega_1, \ldots, \omega_k)$ in $\mathbb{Q}^k$, and k arbitrary, consider the set $\mathcal{G}_\rho = \bigcup_{\omega_i \in \rho} \Gamma_{\omega_i}$, union of GC's whose projections do not intersect. Let

$$J_{[a,b]} = \text{closure}\{(x,y) \in \left(\text{CO}_{[a,b]}\right)^2 \mid I(\tau_{m,n}x, y) \leq 1, \quad \forall (m,n) \in \mathbb{Z}^2\}\,.$$

This is a compact attractor block for the flow on the cartesian product. Let $K \subset J_{[a,b]}$ be the maximum invariant set in $J_{[a,b]}$. Then K and its projection K' on the first component are both compact. Take an increasing (for the inclusion) sequence of finite subsets ρ of $\mathbb{Q}$, say $\{\rho^j\}_{j\in\mathbb{N}}$ such that $\bigcup_{j\in\mathbb{N}} \rho^j = \mathbb{Q} \cap [a,b]$. Since K' is compact, assume that the sequence $\{\mathcal{G}_{\rho^i}\}_{i\in\mathbb{N}}$ converges to a set $\mathcal{L}$ in K'. Now show that for all $\omega \in [a,b]$, $\mathcal{L} \cap \text{CO}_\omega$ contains at least one Ghost Circle. Show that two GCs $\Gamma_\omega, \Gamma_{\omega'}$ of different rotation numbers in $\mathcal{L}$ must satisfy $\Gamma_{\omega'} \cap \Gamma_{\omega'} = \emptyset$. To construct a partition, i.e. a family of non intersecting circles , pick (using the axiom of choice!) one Ghost Circle of $\mathcal{L}$ for each ω in $[a,b]$.

21. Remarks and Applications

A. Remarks

1) Let's note that the techniques introduced in this chapter have a scope that goes beyond proving the vertical ordering of Aubry-Mather sets. Angenent (1988)introduced the flow ζ^t and its monotonicity. He used it to prove, for instance, the existence of periodic orbits that, in the generic case, would come from "elliptic islands around elliptic islands", as well as homoclinic and heteroclinic orbits between hyperbolic points. The remarkable fact is that his results do not make any generic assumption. Indeed, removing generic assumptions about transversality of unstable manifolds is often a major hurdle in proofs that use hyperbolicity, and can be seen as an advantage of variational techniques. As an example, it was this kind of technical hurdle that barred Tangerman & Veerman (1990a)to obtain a complete proof that the Aubry-Mather sets are vertically ordered, a fact that they conjecture in that paper. In a larger context, Angenent (1990) continued exploring the notion of monotonicity and its relationship to the maximum principle of parabolic PDEs and obtained a generalization of the Aubry-Mather theorem. As noted before, Koch & al. (1994) and Candel & de la Llave (1997) use gradient flow technics to prove interesting Aubry-Mather type theorems about functions on lattices of any dimensions. More recently MacKay et al. have ....?

2) Ghost circles first appeared in Golé (1992 a). They stem from an effort I was making in understanding the Ghost Tori of my thesis (see Chapter 4). I had constructed circle within the Ghost Tori. My advisor G. Hall as well as R. MacKay and J. Meiss asked me if I could prove their projections were graphs. This launched the work in Golé (1992 a), where I also recover a result similar to that of

on existence of invariant circles In his work on toral and annulus homeomorphisms, LeCalvez (1997)proposes another way to construct our GCs: take an ordered circle in $\mathrm{CO}_\omega/\mathbb{Z}$ which is $\mathbb{Z}^2$ invariant, but not necessarily ζ^t invariant. A simple choice is the "straight" circle with $x_k(\xi) = k\omega + \xi$. Apply the flow ζ^t to this whole circle, and take a limit as the time $t \to \infty$. Le Calvez suggested to us that letting the flow act on non–intersecting collections of rational GCs may be a way to prove Theorem 18.4. In a way that is reminiscent to Le Calvez' construction of GCs, Fathi (???am) has obtained the generalized Aubry-Mather sets of Mather by applying a flow in a functional analytic space of Lagrangian graphs ???. Finally Katznelson & Ornstein (???am via trimming) find Aubry-Mather sets on a collection of pseudo–graphs that are (not strictly) ordered vertically. They do this by iterating the map on circles in the annulus, trimming the image of the circles at each step so that they remain pseudo–graphs (see Chapter 3). It would be interesting to investigate the parallel between these different methods.

B. Approximate action-angle variables for an arbitrary twist map

Dewar & Meiss (???flux min) attempt the construction of approximate action-angle variables using almost–invariant circles defined through a mean square flux variational principle. We refer the reader to their paper as to the physical relevance of such coordinates. We show here that similar approximate action variables can easily be defined from our GC's. Given any finite number of minimal Aubry-Mather sets, we will produce a continuous foliation of the annulus by circles such that each of the Aubry-Mather set of our chosen collection is contained in a different circle of the foliation. Moreover, such a construction will also produce a completely integrable, albeit not necessarily differentiable map of the annulus that coincides with the original map on the collection of Aubry-Mather sets and leaves the foliation invariant. We sketch here the simple construction.

Let $M_{\omega_1}, \ldots, M_{\omega_n}$ be an arbitrary collection of minimal Aubry-Mather sets. From Theorem 18.1, we know that we can produce a corresponding collection $\Gamma_1, \ldots, \Gamma_n$ of GC's whose disjoint projections contain the chosen Aubry-Mather sets. Parameterize these GC's by their coordinate x_0 as in (16.1) and order them by increasing rotation number. Between two succesive GC's , say Γ_k and Γ_{k+1}, construct the continuous family:

$$\begin{aligned} &\Gamma_t(\xi) = \big(\cdots, x^t_{-1}(\xi), \xi, x^t_1(\xi), \cdots\big) \\ \text{with} \quad &x^t_1(\xi) = (1-t)x^{(k)}_1(\xi) + t x^{(k+1)}_1(\xi) \\ &x^t_j(\xi) = (x^t_1)^j(\xi) \end{aligned}$$

where, since both $x_1^{(k)}$ and $x_1^{(k+1)}$ are lifts of homeomorphisms of the circle, x_1^t also is (it must be periodic and monotone); $(x_1^t)^j$ represents the jth iterate of this homeomorphism. It is not hard to see that, for $t \neq 0$ or 1, Γ_t has all the properties of a GC except for that of being invariant under the flow. In particular it is a circle in $\mathrm{CO}_{\omega_t}/\tau_{0,1}$ on which the shift $\tau_{1,0}$ acts as a circle homeomorphism with rotation number $\omega_t = (1-t)\omega_k + t\omega_{k+1}$. Its projection $\pi\Gamma_t$ is a graph in the annulus. The circles $\pi\Gamma_t$ do not intersect for different t's since in the (x_0, x_1) coordinates, they are the linear interpolation along the x_1 axis of the non intersecting graphs of $x_1^{(k)}$ and $x_1^{(k+1)}$. Repeating this process between each pair of adjacent Γ_k's in our finite collection gives the continous foliation $\pi\Gamma_t$ advertised. The completely integrable map is given by $\tau_{1,0}$ acting on the family Γ_t of Ghost Circles, or alternatively by $\pi \circ \tau_{1,0} \circ \pi^{-1}$ acting on the annulus, which is the topologically embedded image (by π) of the family Γ_t.

Since for generic maps the rational GC's can be made C^1, the above construction yields, when starting with a generic map and rational Aubry- Mather sets, a C^1 foliation (after smoothing the glueing of our interpolations). All the minimizing periodic orbits of the chosen rotation numbers are then embedded in the construction. One can also take a limit of this process, by adding more and more Aubry- Mather sets. One obtains an ordered continuum of circles in $\mathbb{R}^{\mathbb{Z}}$ which contains our set $\mathcal{L}$ of the proof of Theorem 18.1. Alternatively, we could have started with the set $\mathcal{L}$ of GCs and filled its gaps as above, all at once (gaps will occur between the Γ_ω^- and the Γ_ω^- of a given rotation number).

Further study of this object might be interesting in order to draw a parallel between twist maps and families of circle maps, *eg.* in the theory of renormalization (see MacKay (??? wsp book)).

C. Partition for transport

In the theory of transport of MacKay, Meiss and Percival MacKay, Meiss & Percival (1984), MacKay, Meiss & Percival (1986), it is sought to use almost invariant circles in order to form disjoint boxes containing the "resonance zones" around the elliptic islands (or hyperbolic points with reflexion) of the periodic minimax orbits of different rational rotation numbers.

It is not hard to see that the pairs $C_{p/q\pm}$ of projections of the $p/q\pm$ GC's each enclose the circle $C_{p/q}$ of Theorem 18.1: they are defined as limits of circles that are respectively strictly above or strictly below $C_{p/q}$. Moreover, as in the almost invariant circles (or partial separatrices) of MacKay, Meiss & Percival (1986), $C_{p/q}$ and the $C_{p/q\pm}$ all meet at the minimum p/q orbits, at least when there are finitely many of them (i.e. generically). $C_{p/q+}$ (resp. $C_{p/q-}$) contains the advance (resp. retrograde) homoclinic orbits (min and minimax), by an argument of Katok (see Hasselblat & Katok (1995)). We therefore hope that the boxes defined by the pairs $C_{p/q\pm}$ of GC's may be used as intended for the partial separatrices in MacKay, Meiss &

Percival (1986). The advantage of our boxes over those formed by partial separatrices is that their boundaries are graphs and that they are disjoint from one another (statements unproven to our knowledge for partial separatrices in the general case. See Tangerman & Veerman (1990a) for partial results). Hence the calculation of the flux through them does not rely on the hypothesis that the turnstiles of MacKay et al. always have the simple shape of a figure 8. One of the advantages of their partial barriers is that they can canalise the flux through "cheminees", i.e., points exit a resonance zone through one turnstile (as opposed to infinitely many in our case). We refer the reader to MacKay & Muldoon for further arguments in favor of the use of Ghost Circles in transport theory, as well as some very intersesting pictures.

D. An extension of Aubry's Fundamental Lemma

As a consequence of Theorem 18.4, we get that any pairs of points in two unlinked maximal skeletons of distinct rotation numbers have intersection index 1. By Aubry Lemma, we knew this to be the case for minimizers, but our results shows that it is also true for the minimaxes and local minima in the skeletons. The relevance of this appears clearer in the light of LeCalvez (astérisque) , where he shows that this intersection number is in fact a linking number for the corresponding orbits of the suspension flow of the map. Extending an idea of Hall (1984), he shows that this linking is an obstruction to continue periodic orbits simultaneously, through paths of periodic orbits in an isotopy of the map to some completely integrable twist map. In our terminology, his result implies that the periodic orbits corresponding to critical sequences in a minimally linked set can "continue" simultaneously through curves of periodic orbits of an isotopy of our map to a well chosen completely integrable map. In particular, LeCalvez already noted that, because of Aubry's Fundamental Lemma, any collection of minimum periodic orbits can be continued simultaneously to orbits of a completely integrable map. A consequence of Theorem 18.4, where we construct minimally linked sets that contain minimum and minimax orbits, we get, using LeCalvez' result, orbits of minimax type as well as periodic minimizers continuing simultaneously to orbits of a completely integrable map f_0, through paths of periodic orbits of a curve of maps joining f to f_0.

22. Appendix: Proofs of Monotonicity and the Sturmian Lemma

In this appendix, we give the proofs of Theorem 14.2 and Lemma 14.3. Eventhough it is a consequence of the latter, we start with a simpler, direct proof of the former. Both proofs are by S. Angenent.

A. Proof of Strict Monotonicity

We let the reader show that if the operator solution of the linearised equation:

$$\dot{\boldsymbol{u}}(t) = L\boldsymbol{u}(t) \tag{22.1}$$

with

$$L : \{v_k\}_{k\in\mathbb{Z}} \mapsto \{\beta_k v_{k-1} + \alpha_k v_k + \beta_{k+1} v_{k+1}\}_{k\in\mathbb{Z}}$$

$$\alpha_k = -\partial_{22}S(x_{k-1}, x_k) - \partial_{11}S(x_k, x_{k+1}), \quad \beta_k = -\partial_{12}S(x_{k-1}, x_k)$$

is strictly positive, then the flow is strictly monotone. $L(\boldsymbol{x}(t))$ is an infinite tridiagonal matrix with positive off diagonal terms $-\partial_{12}S(x_k, x_{k+1})$ (see 17.1 for a finite dimensional version) . The diagonal terms

$\partial_{11}S(x_k, x_{k+1}) + \partial_{22}\partial_2 S(x_{k-1}, x_k)$ are uniformaly bounded by assumption. Hence, for any $T > 0$ for which $\boldsymbol{x}(t)$ is defined when $0 \leq t \leq T$, we can find a positive λ such that:

$$B(t) = L(\boldsymbol{x}(t)) + \lambda Id$$

is a positive matrix with strictly positive off diagonal terms. If $\boldsymbol{u}(t)$ is solution of the equation (22.1) then $e^{\lambda t}\boldsymbol{u}(t)$ is solution of :

$$\dot{\boldsymbol{v}}(t) = B(t)\boldsymbol{v}(t), \tag{22.2}$$

hence the strict positivity of the solution operator for(22.1) is equivalent to that of (22.2) . Looking at the integral equation:

$$\boldsymbol{v}(t) = \boldsymbol{v}(0) + \int_0^t B(s)\boldsymbol{v}(s)ds,$$

one sees that Picard's iteration will give positive solutions for positive vector $\boldsymbol{v}(0)$. This will imply, assuming that $v_k(0) > 0, v_l(0) \geq 0$, for $l \neq k$:

$$v_{k+1}(t) \geq v_{k+1}(0) + \int_0^t B_{k,k+1}(s)v_k(s)ds > 0$$

The same holding for v_{k-1}. By induction, $v_k(t) > 0, \forall k \in \mathbb{Z}$ and the operator solution is strictly positive. This finishes the proof of Theorem 14.2. □

B. Proof of the Sturmian Lemma

Let $x_i(t)$ $(i_0 \leq i \leq i_1, -T \leq t \leq t)$ be a solution of

$$\frac{dx_i}{dt} = a_i(t)x_{i-1} + b_i(t)x_i(t) + c_i(t)x_{i+1}(t) \quad (i_0 < i < i_1) \tag{22.3}$$

where we assume that the coefficients $a_i(t), b_i(t), c_i(t)$ are continuous and satisfy

$$a_i(t), c_i(t) \geq \delta; \qquad a_i, b_i, c_i \leq M \tag{22.4}$$

for all $-T \leq t \leq T, i_0 < i < i_1$, and for some constants $\delta, M > 0$.

Lemma 22.1 *Assume*

$$x_i(0) \begin{cases} = 0 & \text{for } i_0 < i < i_1 \\ \neq 0 & \text{if } i = i_0 \text{ or } i = i_1. \end{cases}$$

Then the sequence $\{x_{i_0}(t), \ldots, x_{i_1}(t)\}$ *has less sign changes when* $t > 0$ *than when* $t < 0$.

Proof. First a few reductions. Consider

$$y_i(t) = B_i(t)x_i(t)$$

with $b_i(t) = \exp\{-\int_0^t b_i(\tau)d\tau\}$; then

$$\frac{dy_i}{dt} = A_i(t)y_{i-1} + C_i(t)y_{i+1},$$

where

$$A_i(t) =_{\rm def} \frac{B_{i-1}(t)}{B_i(t)} a_i(t), \qquad C_i(t) =_{\rm def} \frac{B_{i+1}(t)}{B_i(t)} c_i(t).$$

In other words, we may assume that the $b_i(t)$ vanish. Note that $\{x_i(t)\}$ and $\{y_i(t)\}$ have the same sign changes.

The coefficients A_i, C_i satisfy

$$\delta e^{-MT} \le A_i(t), C_i(t) \le M e^{+MT} \tag{22.5}$$

By integrating the differential equation for $y_i(t)$ we find that for $i_0 < i < i_1$ one has

$$y_i(t) = \int_0^t \{A_i(\tau) y_{i-1}(\tau) + C_i(\tau) y_{i+1}(\tau)\} d\tau \tag{22.6}$$

Proposition 22.2 *For $i_0 < i < i_1$ one has*

$$y_i(t) = M_i t^{i-i_0} + N_i t^{i_1-i} + o\left(|t|^{i-i_0} + |t|^{i_1-i}\right) \qquad (t \to 0) \tag{22.7}$$

where the constants M_i and N_i are given by

$$M_i = A_i(0) A_{i-1}(0) \cdots A_{i_0+1}(0) \frac{y_{i_0}(0)}{(i-i_0)!},$$

$$N_i = C_i(0) C_{i+1}(0) \cdots C_{i_1-1}(0) \frac{y_{i_1}(0)}{(i_1-i)!}.$$

We shall prove this by induction. The relevant property of the coefficients M_i, N_i is that the M_i have the same sign as $y_{i_0}(0)$, and the N_i have the same sign as $y_{i_1}(0)$. Furthermore, one of the two terms in (22.7) always dominates the other, unless $i - i_0 = i_1 - i$, i.e. unless $i = \frac{i_0+i_1}{2}$; if $i < \frac{i_0+i_1}{2}$ then $y_i(t) = M_i t^{i-i_0} + o\left(t^{i-i_0}\right)$, if $i > \frac{i_0+i_1}{2}$ then $y_i(t) = N_i t^{i_1-i} + o\left(t^{i_1-i}\right)$

Proof. We may assume $i_1 - i_0 \ge 2$. The $y_i(t)$ are continuous, and hence bounded as $t \to 0$. Therefore it follows from (22.6) that $|y_i(t)| \le C|t|$ for $|t| \le T$.

If $i_1 - i_0 = 2$, then the only i with $i_0 < i < i_1$ is $i = i_0 + 1 = i_1 - 1$, and we have

$$\begin{aligned} y_{i_0+1}(t) &= \int_0^t \{A_{i_0+1}(0) y_{i_0}(0) + C_{i_1-1}(0) y_{i_1}(0) + o(1)\} d\tau \\ &= M_{i_0+1} t + N_{i_0-1} t + o(t), \end{aligned}$$

as claimed.

If $i_1 - i_0 > 2$, then $y_{i_0+2}(t) = o(1)$, and (22.6) implies

$$\begin{aligned} y_{i_0+1}(t) &= \int_0^t \{A_{i_0+1}(0) y_{i_0}(0) + o(1)\} d\tau \\ &= M_{i_0+1} y_{i_0}(0) t + o(t). \end{aligned}$$

Likewise (22.6) implies $y_{i_1-1}(t) = N_{i_0-1} y_{i_1}(0) t + o(t)$. If $i_1 - i_0 = 3$ this proves the claim; if $i_1 - i_0 > 3$, then for all $i_0 + 1 < i < i_1 - 1$ one deduces from (22.6) and the estimate $|y_{i\pm 1}(t)| \le C|t|$ that $|y_i(t)| \le Ct^2$.

The general induction step in the derivation of (22.7) is as follows. Assume that it has been shown that (22.7) holds for all i with $i_0 < i < i_0 + k$, or $i_1 - k < i < i_1$; moreover assume it has been shown that $|y_i(t)| \le C|t|^k$ for $i_0 + k \le i \le i_1 - k$.

If $i_0+k=i_1-k$, then (22.7) implies

$$y_{i_0+k}(t)=\int_0^t\{A_{i_0+k}(0)M_{i_0+k-1}\tau^{k-1}+C_{i_1-k}(0)N_{i_1-k+1}\tau^{k-1}+o\left(\tau^{k-1}\right)\}d\tau$$
$$=M_{i_0+k}t^k+N_{i_1-k}t^k+o\left(t^k\right),$$

with

$$M_{i_0+k}=A_{i_0+k}(0)\frac{1}{k}M_{i_0+k-1},$$
$$N_{i_1-k}=C_{i_1-k}(0)\frac{1}{k}N_{i_1-k+1}.$$

In this case the claim is proved.

Otherwise $i_0+k<i_1-k$, and a similar computation shows that (22.7) holds when $i=i_0+k$ and $i=i_1-k$. Finally, using (22.6) again, one finds that for $i_0+k<i<i_1-k$ the estimate $\left|y^{i\pm1}(t)\right|\le C\,|t|^k$ implies $|y_i(t)|\le C\,|t|^{k+1}$, which completes the induction step.

Lemma 22.1follows directly from the proposition. If $y_{i_0}(0)$ and $y_{i_1}(0)$ have the same sign, say they are positive, then the expansion (22.7) implies that all $y_i(t)$ are positive for $t>0$; For small negative t the sequence $y_{i_0}(t),y_{i_0+1}(t),\ldots,y_{i_1}(t)$ alternates signs, except in the middle, i.e. if i_1-i_0 is odd then $y_{i_0+k}(t)$ and $y_{i_0+k+1}(t)$ (with $k=\left[\frac{i_1-i_0}{2}\right]$) will have the same sign.

Indeed, (22.7) says the sequence $\{y_{i_0}(t),\ldots,y_{i_1}(t)\}$ has the signs as the sequence

$$(c_0,c_1t,c_2t^2,\ldots,c_{k-1}t^k,c_kt^k,c_{k+1}t^{k-1},\ldots,c_{2k-1}t,c_{2k})$$

if $i_1-i_0=2k$ is even, and $\{y_{i_0}(t),\ldots,y_{i_1}(t)\}$ will have the same signs as the sequence

$$(c_0,c_1t,c_2t^2,\ldots,c_kt^{k+1},c_{k+1}t^k,\ldots,c_{2k}t,c_{2k+1})$$

if $i_1-i_0=2k+1$ is odd; here the c_j's are positive constants, with the possible exception of the coefficient c_k of t^{k+1} in the second sequence.

If $y_{i_0}(0)$ and $y_{i_1}(0)$ have opposite signs, then one can again use the expansion (22.7) to derive that the sequence $\{y_i(t)\}$ has exactly one sign change for $t>0$, and i_1-i_0-1 sign changes for $t<0$. If $i_1-i_0=2$, then $\{y_{i_0}(t),y_{i_0+1}(t),y_{i_0+2}(t)\}$ is "transverse" to the zero sequence for all small t, whatever the sign of $y_{i_0+1}(t)$ is.

Thus, if $\{y_{i_0}(t),\ldots,y_{i_1}(t)\}$ is *not* transverse to the zero sequence at $t=0$, then either $i_1>i_0+2$, or $i_1=i_0+2$, and $y_{i_0}(0)$ and $y_{i_1}(0)$ have the same sign. In either case we have shown that the number of sign changes of $\{y_{i_0}(t),\ldots,y_{i_1}(t)\}$ drops at $t=0$. □

Lemma 22.1implies the following:

Lemma 22.3 *If $\{x_{i_0}(t),\ldots,x_{i_1}(t)\}$ is a C^1 solution of (22..1), with $x_{i_0}(t),x_{i_1}(t)\neq0$ for all $t_0<t<t_1$, then*
(a) the number of signchanges of $\{x_{i_0}(t),\ldots,x_{i_1}(t)\}$ does not increase;
(b) this number drops whenever $\{x_{i_0}(t),\ldots,x_{i_1}(t)\}$ is not transverse to the zero sequence.

Lemma 22.1also implies the fundamental theorem on intersections which we use in the paper.

Theorem 22.4 *Let $\boldsymbol{x}(\cdot), \boldsymbol{y}(\cdot) \in CO$ be different solutions of*

$$\frac{dx_k}{dt} = -\partial_2 S(x_{k-1}, x_k) - \partial_1 S(x_k, x_{k+1}) \ ;$$

then $I\left(\boldsymbol{x}(t), \boldsymbol{y}(t)\right)$ does not increase, and decreases whenever $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ are not transverse.

Proof. By the mean value theorem the difference $\boldsymbol{z}(t) = \boldsymbol{x}(t) - \boldsymbol{y}(t)$ satisfies a linear equation of the form (22.1). If $\boldsymbol{x}(t_0) \pitchfork \boldsymbol{y}(t_0)$, then $I\left(\boldsymbol{x}(t), \boldsymbol{y}(t)\right)$ is constant for t near t_0.

If $\boldsymbol{x}(t_0)$ and $\boldsymbol{y}(t_0)$ are not transverse, then since $\boldsymbol{x}(t_0) \neq \boldsymbol{y}(t_0)$ one can choose $i_0 < i_1$ such that $z^{i_0}(t_0) \neq 0$, $z^{i_1}(t_0) \neq 0$, while $z^i(t_0) = 0$ for $i_0 \leq i \leq i_1$. Lemma 22.1then implies the theorem. □

The other situation in which we use the result about $I\big(\boldsymbol{x}(t), \boldsymbol{y}(t)\big)$, i.e. the case when $\boldsymbol{x}(t), \boldsymbol{y}(t)$ belong to different X_ω's can be dealt with in the same way.

Lemma GTClemmaxskel is 17.3, Lemma GClemgraphordering is 16.4, Lemma GClemmonlimgc is 20.2, Theorem GCthmamordered is 18.1, Theorem GCthmdisjointgc is 18.4, Lemma GClemmountainpass is 17.4, Theorem 15.1is GCthmamflowproof, Corollary GCcorocoinv is 14.4, Theorem GCthmmonotoneflow is 14.2

Chapter 4 or STM

SYMPLECTIC TWIST MAPS

9/25/99

This is the version revised on January 12 1998. It minimizes the use of symplectic theory or of homotopy (in the torus case). The general case could be moved to another chapter. Birkhoff-Lewis: point to the idea of proof: intersection of Lag tori. Get the hyperbolic metric right. Find page in Gallot on diffeo $TM \cong M$. Find the whereabouts of Eduardo's picture (which map, which orbit). State Birkhoff's normal form for invariant diophantine tori (see Yoccoz, page 754-07, Herman IMA?)

In this chapter, we generalize the definition of twist maps of the annulus to that of symplectic twist maps in higher dimensions. In many cases, around elliptic fixed points, area preserving planar maps yield twist maps of the annulus $\mathbb{S}^1 \times \mathbb{R}$. Likewise, symplectic maps in $\mathbb{R}2n$ around their elliptic fixed points lead to symplectic twist maps of $\mathbb{T}^n \times \mathbb{R}^n$, the cotangent bundle of the n dimensional torus. This is one among many other reasons which make $\mathbb{T}^n \times \mathbb{R}^n$ one of the most natural spaces to study. Another reason is that, although these notions are at least implicitly present, almost no knowledge of manifolds, fiber bundles and differential forms is needed for the study of symplectic maps on this space. Hence we devote the first sections of this chapter to defining symplectic twist map of $\mathbb{T}^n \times \mathbb{R}^n$ and exploring their relationship with their generating functions.

Nonetheless, cotangent bundles of many other manifolds do occur in mechanics (*eg.* the configuration space of the solid rigid body is SO(3)) and there too it is possible to define and make use of symplectic twist maps. For this part of the chapter, the reader should be familiar with the notion of cotangent bundle, differential forms as are given in Section 46 of Appendix 1 or SG.

23. Symplectic Twist Maps of $\mathbb{T}^n \times \mathbb{R}^n$

A. Definition

Let $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ be the n–dimensional torus. An analog to the annulus in higher dimensions which is most natural in mechanics is the space $\mathbb{T}^n \times \mathbb{R}^n$, which can be seen as the cartesian product of n annuli. We give $\mathbb{T}^n \times \mathbb{R}^n$ the coordinate $(\boldsymbol{q}, \boldsymbol{p}) = (q_1, \dots, q_n, p_1, \dots, p_n)$. In mechanics, $q_1, \dots, q_n$ would be n angular configuration variables of the system, whereas $p_1, \dots, p_n$ would be their conjugate momentum, and $\mathbb{T}^n \times \mathbb{R}^n$ is the cotangent bundle $T^*\mathbb{T}^n$ of the torus $\mathbb{T}^n$.

The following is a generalization of the definition of twist maps of the cylinder:

Definition 23.1 Let F be a diffeomorphism of $\mathbb{R}2n$ and write $(\boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p}),\boldsymbol{P}(\boldsymbol{q},\boldsymbol{p})) = F(\boldsymbol{q},\boldsymbol{p})$. Let F satisfies:

1) $F(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{p}) = F(\boldsymbol{q},\boldsymbol{p}) + (\boldsymbol{m},0)$
2) *Twist Condition:* the map $\psi_F : (\boldsymbol{q},\boldsymbol{p}) \mapsto (\boldsymbol{q},\boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p}))$ is a diffeomorphism of $\mathbb{R}2n$.
3) *Exact Symplectic:* In the coordinates $(\boldsymbol{q},\boldsymbol{Q})$,

$$\boldsymbol{P}d\boldsymbol{Q} - \boldsymbol{p}d\boldsymbol{q} = dS(\boldsymbol{q},\boldsymbol{Q}) \tag{23.1}$$

where S is a real valued function on $\mathbb{R}2n$ satisfying:

$$S(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{Q}+\boldsymbol{m}) = S(\boldsymbol{q},\boldsymbol{Q}), \quad \forall \boldsymbol{m} \in \mathbb{Z}^n. \tag{23.2}$$

Then the map f that F induces on $\mathbb{T}^n \times \mathbb{R}^n$ is called a *Symplectic Twist Map*.

As for maps of the annulus, $S(\boldsymbol{q},\boldsymbol{Q})$ is called a *generating function* of the map F: Equation (23.1) is equivalent to

$$\begin{aligned} \boldsymbol{p} &= -\,\partial_1 S(\boldsymbol{q},\boldsymbol{Q})) \\ \boldsymbol{P} &= \partial_2 S(\boldsymbol{q},\boldsymbol{Q}), \end{aligned} \tag{23.3}$$

and thus F is implicitely given by S since

$$\begin{aligned} F(\boldsymbol{q},\boldsymbol{p}) &= (\boldsymbol{Q}\circ\psi_F(\boldsymbol{q},\boldsymbol{p}), \partial_2 S\circ\psi_F(\boldsymbol{q},\boldsymbol{p})) \quad \text{with} \\ \psi_F^{-1}(\boldsymbol{q},\boldsymbol{Q}) &= (\boldsymbol{q}, -\partial_1 S(\boldsymbol{q},\boldsymbol{Q})) \end{aligned} \tag{23.4}$$

Note that the prescription of F through its generating function S is often more theoretical than computational: it involves the inversion of the diffeomorphism ψ_F^{-1}.

B. Comments on the Definition

(1) The periodicity condition $F(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{p}) = F(\boldsymbol{q},\boldsymbol{p})+(\boldsymbol{m},0)$ implies that F induces a map f on $\mathbb{T}^n\times\mathbb{R}^n$. It also implies that (in fact is equivalent to) f is homotopic to Id (see the Exercise 23.1).

(2) The twist condition (2) of definition 23.0 implies the local twist condition often used in the litterature:

$$\text{Condition}(2') \qquad \det\ \partial\boldsymbol{Q}/\partial\boldsymbol{p} \neq 0,$$

We will explore in Section 26extra assumptions under which the local twist implies the global twist of Condition (2).

(3) In terms of differential forms, $\boldsymbol{P}d\boldsymbol{Q} - \boldsymbol{p}d\boldsymbol{q} = F^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q}$. The periodicity of S given by $S(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{Q}+\boldsymbol{m}) = S(\boldsymbol{q},\boldsymbol{Q})$ in the $(\boldsymbol{q},\boldsymbol{Q})$ coordinates becomes $S(\boldsymbol{q}+m,\boldsymbol{p}) = S(\boldsymbol{q},\boldsymbol{p})$ in the $(\boldsymbol{q},\boldsymbol{p})$ coordinates ($i.e.$ applying Ψ_F^{-1}). In particular S induces a function s on $\mathbb{T}^n\times\mathbb{R}^n$ such that $f^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = ds$ ($\boldsymbol{q}$ is seen as coordinate on $\mathbb{T}^n$ here). This last equality expresses the fact that f is *exact symplectic*. As is made more precise in Chapter SG, if f is exact symplectic it is also *symplectic*:

$$f^*\,pd\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = ds \Rightarrow d(f^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q}) = 0 \Rightarrow f^*d\boldsymbol{p}\wedge d\boldsymbol{q} = d\boldsymbol{p}\wedge d\boldsymbol{q}.$$

Any symplectic map of $\mathbb{R}2n$ is exact symplectic, but it is not true of maps of $\mathbb{T}^n\times\mathbb{R}^n$: the map $f(\boldsymbol{q},\boldsymbol{p}) \mapsto (\boldsymbol{q},\boldsymbol{p}+\boldsymbol{m}), \boldsymbol{m}\neq 0$ is symplectic but not exact symplectic. As for maps of the annulus, exact symplecticity can be interpreted as a zero flux condition, but the flux is now an n dimensional quantity.

Exercise 23.2 Each homeomorphism of the torus $\mathbb{T}^n$ is homotopic to a unique torus map induced by a linear map A of $Gl(n, \mathbb{Z})$ (the group of invertible integer $n \times n$ matrices). Likewise, each homotopy classes of homeomorphisms of $\mathbb{T}^n \times \mathbb{R}^n$ has exactly one representant of the form $A \times Id$ where $A \in Gl(n, \mathbb{Z})$. Show that any lift F of a map homotopic to $A \times Id$ satisfies:

$$F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P}) \Rightarrow F(\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{p}) = (\boldsymbol{Q} + A\boldsymbol{m}, \boldsymbol{P})$$

.

Exercise 23.3 Show that if $F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})$ is the lift of a symplectic twist map with generating function $S(\boldsymbol{q}, \boldsymbol{Q})$, then $F^{-1}(\boldsymbol{Q}, \boldsymbol{P}) = (\boldsymbol{q}, \boldsymbol{p})$ is also the lift of a symplectic twist map with generating function $-S(\boldsymbol{Q}, \boldsymbol{q})$.

Exercise 23.4 Show that if F and F' are two lifts of the same symplectic twist map F, their corresponding generating functions S and S' satisfy:

$$S(\boldsymbol{q}, \boldsymbol{Q}) = S'(\boldsymbol{q}, \boldsymbol{Q} + \boldsymbol{m}),$$

where $\boldsymbol{m} \in \mathbb{Z}^n$ is such that $F' = T_{\boldsymbol{m}} \circ F$.

C. The Variational Setting

As in the case of monotone twist maps of the annulus, the generating function of a symplectic twist map induces a variational approach to finding orbits of the map.

Proposition 24.1 (Critical Action Principle) *Let $f_1, \dots, f_N$ be symplectic twist maps of $T^*\mathbb{T}^n$, and let F_k be a lift of F_k, with generating function S_k. There is a one to one correspondence between orbits segments $\{(\boldsymbol{q}_{k+1}, \boldsymbol{p}_{k+1}) = F_k(\boldsymbol{q}_k, \boldsymbol{p}_k)\}$ under the successive F_k's and the sequences $\{\boldsymbol{q}_k\}_{k\in\mathbb{Z}}$ in $(\mathbb{R}^n)^{\mathbb{Z}}$ satisfying:*

$$\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) + \partial_2 S_{k-1}(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k) = 0 \tag{24.1}$$

The correspondence is given by: $\boldsymbol{p}_k = -\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$.

Proof. It is identical to the case $n = 1$, Corollary 5.2. □

As in the case $n = 1$, Equation (24.1) can be interpreted as:

$$\nabla W(\overline{\boldsymbol{q}}) = 0 \quad \text{with}$$
$$W(\overline{\boldsymbol{q}}) = \sum_0^{N-1} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}).$$

25. Examples

Example 25.1 The Generalized Standard Map

The *generalized standard map* or *standard family* is the family of symplectic twist map whose lift is generated by the following functions:

$$S_\lambda(\boldsymbol{q},\boldsymbol{Q}) = \frac{1}{2}\left\|\boldsymbol{Q}-\boldsymbol{q}\right\|^2 + V_\lambda(\boldsymbol{q}).$$

where V_λ is a family of C^2 functions that are $\mathbb{Z}^n$–periodic, λ a parameter on some euclidian space and $V_0 \equiv 0$. It is trivial to see that S satisfies the periodicity condition $S_\lambda(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{Q}+\boldsymbol{m}) = S_\lambda(\boldsymbol{q},\boldsymbol{Q})$. To find the corresponding map, we compute:

$$\begin{aligned}\boldsymbol{p} &= -\partial_1 S_\lambda(\boldsymbol{q},\boldsymbol{Q}) = \boldsymbol{Q}-\boldsymbol{q}-\nabla V_\lambda(\boldsymbol{q})\\ \boldsymbol{P} &= \partial_2 S_\lambda(\boldsymbol{q},\boldsymbol{Q}) = \boldsymbol{Q}-\boldsymbol{q}\end{aligned}$$

from which we immediately get:

$$\begin{aligned}\boldsymbol{Q} &= \boldsymbol{q}+\boldsymbol{p}+\nabla V_\lambda(\boldsymbol{q})\\ \boldsymbol{P} &= \boldsymbol{p}+\nabla V_\lambda(\boldsymbol{q})\end{aligned}$$

In other words, the standard map is given by:

$$F_\lambda(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}+\boldsymbol{p}+\nabla V_\lambda(\boldsymbol{q}), \boldsymbol{p}+\nabla V_\lambda(\boldsymbol{q})). \tag{25.1}$$

In the case $n=2$, the following is the most widely studied potential. It is due to Froeschlé (1972)(see also Kook & Meiss (1989), Froeschlé & Laskar (199???)):

$$V_\lambda(q_1,q_2) = \frac{1}{(2\pi)^2}\{K_1\cos(2\pi q_1) + K_2\cos(2\pi q_2) + h\cos(2\pi(q_1+q_2))\}.$$

In this case $\lambda = (K_1, K_2, h) \in \mathbb{R}^3$, and the standard family attached to this potential is a three parameter family of symplectic maps of $\mathbb{T}^2 \times \mathbb{R}^2$. The picture on the bookcover represents the stable and unstable manifolds of a periodic orbit ??? for this map, with parameter???.

When $\lambda = 0$, the map F_λ of (25.1) becomes:

$$F_0(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}+\boldsymbol{p},\boldsymbol{p}).$$

This is an instance of a *completely integrable* symplectic twist map: such maps preserve a foliation of $\mathbb{T}^n\times\mathbb{R}^n$ by tori homotopic to $\mathbb{T}^n \times \{0\}$. On the covering space of each of these tori, the lift of the map is conjugated to a rigid translation. The term "completely integrable" comes from the corresponding notion in Hamiltonian systems (see Example 25.3.)

The reason why the standard map has attracted so much research is that it is a *computable* example in which one may try to understand questions about persistence of invariant tori as the parameter λ varies away from 0, as well as study the various properties of its periodic orbits.

Examples 25.4 Hamiltonian systems

Historically, symplectic twist map appeared as Poincaré return maps in Hamiltonian systems. We develop this idea in Section 19.

Hamiltonian systems in $T^*\mathbb{T}^n$ have also another way of yielding symplectic twist maps: when restricted to an appropriate domain, the time ϵ map of a Hamiltonian system is often a symplectic twist maps.

As a basic example, the Hamiltonian flow generated by:

$$H(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\langle A\boldsymbol{p},\boldsymbol{p}\rangle \quad \text{with} \quad A^t = A,\ \det A \neq 0$$

is completely integrable, in that it preserves each torus $\{\boldsymbol{p}=\boldsymbol{p}_0\}$ and its time t map:

$$g^t(\boldsymbol{q},\boldsymbol{p})=(\boldsymbol{q}+t(A\boldsymbol{p}),\boldsymbol{p})$$

is a completely integrable symplectic twist map. If A is positive definite, g^t restricted to $\{H=1\}$ is just the geodesic flow for the flat metric $\frac{1}{2}\langle A^{-1}\boldsymbol{v},\boldsymbol{v}\rangle$ on $\mathbb{T}^n$. (See 26.)

More generally, if $F(\boldsymbol{q},\boldsymbol{p})=(\boldsymbol{Q},\boldsymbol{P})$ is the lift of the time ε of some Hamiltonian function H, then:

$$\begin{aligned} Q &= q(\epsilon)=q(0)+\epsilon.H_p+o(\epsilon^2)\\ P &= p(\epsilon)=p(0)-\epsilon.H_q+o(\epsilon^2), \end{aligned}$$

and F satisfies the local twist condition "$\frac{\partial Q}{\partial p}(z(0))$ is non degenerate" whenever H_{pp} is non degenerate. This remark was made by Moser (1986) in the dimension 2 case. From this local argument we will derive conditions under which the time ε of a Hamiltonian is a symplectic twist map .

We will also see that, even if the time ϵ map of a Hamiltonian system is not twist, its time 1 map can, for large classes of Hamiltonian systems, still be decomposed into the product of twist maps. Chapter 4 explores these issues in detail.

Exercise 25.5 Compute the expression of the lift of a symplectic twist map generated by:

$$S(\boldsymbol{q},\boldsymbol{Q})=\frac{1}{2}\langle A(\boldsymbol{Q}-\boldsymbol{q}),(\boldsymbol{Q}-\boldsymbol{q})\rangle+\boldsymbol{c}.(\boldsymbol{Q}-\boldsymbol{q})+V(\boldsymbol{q}).$$

Where A is a nondegenerate $n\times n$ symmetric matrix. (This is yet a further generalization of the standard map.)

26. More On Generating Functions

In this section, we explore more in detail the relationship between generating functions and symplectic twist maps.

Proposition 26.1 *There is a homeomorphism*[4] *between the set of lifts F of C^1 symplectic twist maps of $T^*\mathbb{T}^n$ and the set of C^2 real valued functions S on $\mathbb{R}2n$ satisfying the following:*

(a) $S(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{Q}+\boldsymbol{m})=S(\boldsymbol{q},\boldsymbol{Q}),\quad \forall \boldsymbol{m}\in\mathbb{Z}^n$

(b) The maps: $\boldsymbol{q}\to\partial_2 S(\boldsymbol{q},\boldsymbol{Q}_0)$ and $\boldsymbol{Q}\to\partial_1 S(\boldsymbol{q}_0,\boldsymbol{Q})$ are diffeomorphisms of $\mathbb{R}^n$ for any $\boldsymbol{Q}_0$ and $\boldsymbol{q}_0$ respectively.

(c) $S(0,0)=0$.

This correspondence is given by:

$$F(\boldsymbol{q},\boldsymbol{p})=(\boldsymbol{Q},\boldsymbol{P})\Leftrightarrow\begin{cases}\boldsymbol{p}= & -\partial_1 S(\boldsymbol{q},\boldsymbol{Q}))\\ \boldsymbol{P}= & \partial_2 S(\boldsymbol{q},\boldsymbol{Q})).\end{cases}\tag{26.1}$$

Proof. Let F be a lift of a symplectic twist map and $S(\boldsymbol{q},\boldsymbol{Q})$ be its generating function. For such F and S, we have already derived (26.1) from $\boldsymbol{P}d\boldsymbol{Q}-\boldsymbol{p}d\boldsymbol{q}=dS$, and (a) is part of our definition of symplectic twist

[4] In the compact open topologies of the corresponding sets

maps . To show that S satisfies (b), first notice that, by (26.0) , $\boldsymbol{Q} \to -\partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q})$ is just the inverse of the map $\boldsymbol{p} \to \boldsymbol{Q}(\boldsymbol{q}_0, \boldsymbol{p})$, which is a diffeomorphism since $\psi_F : (\boldsymbol{q}, \boldsymbol{p}) \mapsto (\boldsymbol{q}, \boldsymbol{Q})$ is a diffeomorphism by the twist condition. We also have the composition of diffeomorphisms:

$$(\boldsymbol{q}, \boldsymbol{Q}) \stackrel{\psi_F^{-1}}{\to} (\boldsymbol{q}, \boldsymbol{p}) \stackrel{F}{\to} (\boldsymbol{Q}, \boldsymbol{P})$$

which implies that the map $\boldsymbol{q} \to \boldsymbol{P}(\boldsymbol{q}, \boldsymbol{p}_0)$ is a diffeomorphism (that is, F^{-1} satisfies the twist condition), which finishes to prove that S satisfies (b). Since two generating functions of the same F only differ by a constant there is exactly one such $S(0,0) = 0$.

Conversely, given an S satisfying (b), we can define a C^1 exact symplectic map F of $\mathbb{R}2n$ by:

$$(26.2) \qquad \begin{aligned} F(\boldsymbol{q}, \boldsymbol{p}) &= (\boldsymbol{Q} \circ \psi_F(\boldsymbol{q}, \boldsymbol{p}), \partial_2 S \circ \psi_F(\boldsymbol{q}, \boldsymbol{p})) \\ \text{where} \quad \psi_F^{-1}(\boldsymbol{q}, \boldsymbol{Q}) &= (\boldsymbol{q}, -\partial_1 S(\boldsymbol{q}, \boldsymbol{Q})). \end{aligned}$$

It is easy to check that such a pair F, S satisfies (26.1) . Since S satisfies (a), F a lift of a diffeomorphism of $T^*\mathbb{T}^n$: (a) also holds for $\partial_1 S$ and $\partial_2 S$, which implies that $F(\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{p}) = (\boldsymbol{Q} + \boldsymbol{m}, \boldsymbol{P})$ whenever $F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})$. Exercise 23.2shows that F must be homotopic to the Identity. Because of (b), F satisfies the twist condition. Hence the map F (uniquely) defined from (26.1) is a symplectic twist map and it is not hard to see that the correspondence we built between the maps F and the functions S is continuous in the C^1 and C^2 compact open topologies respectively. □

In practice, to recognize whether a function S on $\mathbb{R}2n$ is a generating function for some F, it is usefull to have a criterion to decide when S satisfies condition (b) in Proposition 26.0. This is the purpose of the following Propositions:

Proposition 26.2 *Let* $S: \quad \mathbb{R}2n \to \mathbb{R}$ *be a* C^2 *function satisfying:*

$$(26.3) \qquad \begin{aligned} &(i) S(\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{Q} + \boldsymbol{m}) = S(\boldsymbol{q}, \boldsymbol{Q}), \quad \forall \boldsymbol{m} \in \mathbb{Z}^n \\ &(ii) \det\, \partial_{12} S \neq 0 \\ &(iii) \sup_{(\boldsymbol{q}, \boldsymbol{Q}) \in \mathbb{R}2n} \left\| (\partial_{12} S(\boldsymbol{q}, \boldsymbol{Q}))^{-1} \right\| = K < \infty. \end{aligned}$$

Then S *is the generating function for the lift of a symplectic twist map .*

Proof. The proof is an immediate consequence of Lemma 26.3 applied to the two maps $\boldsymbol{q} \to \partial_2 S(\boldsymbol{q}, \boldsymbol{Q}_0)$ and $\boldsymbol{Q} \to \partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q})$ (note that $\|(\partial_{21} S)^{-1}\| = \|(\partial_{12} S)^{-1}\|$) and of Proposition 26.1.

Lemma 26.3 *Let* $f: \mathbb{R}^N \to \mathbb{R}^N$ *be a local diffeomorphism at each point, such that:*

$$\sup_{x \in \mathbb{R}^N} \left\| (Df_x)^{-1} \right\| = K < \infty.$$

Then f *is a global diffeomorphism .*

We postpone the proof of this lemma to the end of the section.

□

The following Proposition gives a condition under which the local twist condition can be made global.

Proposition 26.4 *Let $F(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{Q},\boldsymbol{P})$ be a symplectic map of* $\mathbb{R}2n$ *with $F(\boldsymbol{q}+\boldsymbol{m},\boldsymbol{p}) = (\boldsymbol{Q}+\boldsymbol{m},\boldsymbol{P})$. Suppose that*

$$\sup_{(\boldsymbol{q},\boldsymbol{p})\in\mathbb{R}2n} \left\|(\partial \boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p})/\partial \boldsymbol{p})^{-1}\right\| < \infty. \tag{26.4}$$

Then F is the lift of a symplectic twist map .

Proof. By Lemma 26.3, for each fixed $\boldsymbol{q}$, the map $\boldsymbol{p} \to \boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p})$ is a global diffeomorphism of $\mathbb{R}^n$. This implies that $\psi_F : (\boldsymbol{q},\boldsymbol{p}) \to (\boldsymbol{q},\boldsymbol{Q})$ is a global diffeomorphism of $\mathbb{R}2n$. □

Proof of Lemma 26.0 We first prove that f is onto. Let $y_0 = f(0)$ and take any $y \in \mathbb{R}^N$. Let $y(t) = (1-t)y_0 + ty$. By the inverse function theorem, f^{-1} is defined and differentiable on an interval $y([0,\epsilon))$. Let a be the supremum of all such ϵ in $[0,1]$. If we prove that f^{-1} is also defined and differentiable at a, then $a = 1$, otherwise, by the inverse function theorem, we get the contradiction that f^{-1} is defined on $[0, a+\alpha)$, for some $\alpha > 0$. For any $t_0, t_1 \in [0,a)$, we have:

$$\begin{aligned} \left\|f^{-1}(y(t_1)) - f^{-1}(y(t_0))\right\| &\leq \sup_{t\in[0,a)} \left\|Df^{-1}(y(t))\right\| \|y - y_0\| \, |t_1 - t_0| \\ &\leq K \|y - y_0\| \, |t_1 - t_0| \, . \end{aligned}$$

So that, for any sequence $t_k \to a$, the sequence $f^{-1}(y(t_k))$ is Cauchy. This proves the existence of $f^{-1}(y(a))$, which implies that f is onto. Since f is onto and open, it is a covering map from $\mathbb{R}^N$ to $\mathbb{R}^N$. Such a covering has to be one sheeted, since $\mathbb{R}^N$ is connected and simply connected. (See Appendix Covering spaces.) This finishes the proof. □

Finally, we end this section with a useful formula.

Proposition 26.5 *The following formula relates the differential of a symplectic twist map F to the second derivatives of its generating function:*

$$DF_{(\boldsymbol{q},\boldsymbol{p})} = \begin{pmatrix} -\partial_{11}S.(\partial_{12}S)^{-1} & -(\partial_{12}S)^{-1} \\ \partial_{21}S - \partial_{22}S.\partial_{11}S.(\partial_{12}S)^{-1} & -\partial_{22}S.(\partial_{12}S)^{-1} \end{pmatrix}.$$

where all the partial derivatives are taken at the point $(\boldsymbol{q},\boldsymbol{Q}) = \psi_F(\boldsymbol{q},\boldsymbol{p})$.

Proof. We will show that $\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} = -(\partial_{12}S)^{-1}(\boldsymbol{q},\boldsymbol{Q})$, where, as usual, we have set $F(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{Q},\boldsymbol{P})$. Differentiating the equality: $\boldsymbol{p} = -\partial_1 S(\boldsymbol{q},\boldsymbol{Q})$ with respect to $\boldsymbol{p}$, viewing $\boldsymbol{Q}$ as a function of $\boldsymbol{q},\boldsymbol{p}$, one gets:

$$Id = -\partial_{12}S(\boldsymbol{q},\boldsymbol{Q}).\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}.$$

The computations for the other terms are similar. □

Exercise 26.6 a) Show that if instead of Condition (1) in the definition of symplectic twist maps we ask F to be homotopic to $A \times Id$, where a lift $\tilde{A}$ of A is in $Gl^+(n,\mathbb{Z})$, then Proposition 26.5 remains true, replacing (a) by:

$$S\left(\boldsymbol{q}+\boldsymbol{m}, \boldsymbol{Q}+\tilde{A}(\boldsymbol{m})\right) = S(\boldsymbol{q},\boldsymbol{Q}).$$

b) Find the map generated by

$$S(\boldsymbol{q},\boldsymbol{Q}) = \frac{1}{2}(\boldsymbol{q} - \tilde{A}^{-1}\boldsymbol{Q})^2 + V(\boldsymbol{q})$$

Note that this exercise shows, in particular, that there are plenty of examples of exact symplectic maps of $T^*\mathbb{T}^n$ that are not homotopic to Id and hence cannot be Hamiltonian maps.

Exercise 26.7 Let $\mathbb{B}^n$ denote a *compact* ball in $\mathbb{R}^n$. Show that if $f : \mathbb{B}^n \to \mathbb{R}^n$ is a differentiable map satisfying :

$$\inf_{x \in \mathbb{B}^n} \langle df_x \boldsymbol{v}, \boldsymbol{v}\rangle \geq a\langle \boldsymbol{v}, \boldsymbol{v}\rangle, \quad \forall \boldsymbol{v} \in \mathbb{R}^n$$

then f is an embedding (diffeomorphism on its image) of $\mathbb{B}^n$ in $\mathbb{R}^n$.

27. Symplectic Twist Maps on Cotangent Bundles of General Compact Manifolds

If the manifold M is not covered (topologically) by $\mathbb{R}^n$, problems occur when we want to make the definition of symplectic twist maps of T^*M as global as in $T^*\mathbb{T}^n$: there cannot be a global diffeomorphism from a fiber of T^*M to the universal cover $\tilde{M}$. This is why we must restrict ourselves to a neighborhood U of the 0–section in T^*M, feeling free to take $U = T^*M$ whenever possible.

In the following U will denote an open subset of T^*M such that:

$$\pi^{-1}(\boldsymbol{q}) \cap U \simeq interior(\mathbb{B}^n) \tag{27.1}$$

where $\pi : T^*M \to M$ is the canonical projection, and $\mathbb{B}^n \subset \mathbb{R}^n$ denotes the n-ball. Hence U is a ball bundle over M, diffeomorphic to T^*M, but relatively compact in T^*M. In practice , the neighborhood on which we let our maps act will be of the form:

$$U = \{(\boldsymbol{q},\boldsymbol{p}) \in T^*\tilde{M} \mid H(\boldsymbol{q},\boldsymbol{p}) < K\}$$

for some function H convex in $\boldsymbol{p}$. When it makes sense, we can let $U = T^*M$ or $U = T^*\tilde{M}$ (e.g., when M is covered by $\mathbb{R}^n$). As in Appendix 1 or SG, we denote by λ the canonical one form on T^*M.

Definition 27.1 A *symplectic twist map* F is a diffeomorphism of an open ball bundle $U \subset T^*M$ (as in (27.1)) onto itself satisfying the following:

(1) F is homotopic to Id.
(2) F is *exact symplectic*: $F^*\lambda - \lambda = \underline{S}$ for some real function valued $\underline{S}$ on U.
(3) (*Twist condition*:) the map $\psi_F : U \to M \times M$ given by $\psi_F(\boldsymbol{z}) = (\pi(\boldsymbol{z}), \pi \circ F(\boldsymbol{z}))$ is an embedding.

The function $S = \underline{S} \circ \psi_F^{-1}$ on $\psi_F(U)$ is called the **generating function** for F.

We leave the reader to check that, in coordinates, this is an obvious generalization of the definition of symplectic twist map of $T^*\mathbb{T}^n$, with the appropriate restrictions of domains. If $\tilde{M} \cong \mathbb{R}^n$, one can take $U = T^*M$ and modify the above definition slightly to make it more global by changing (2) into:

(2') If $\tilde{F} : T^*\tilde{M} \to T^*\tilde{M}$ is a lift of F, the map $\psi_{\tilde{F}} : \tilde{U} \to M \times M$ given by $\psi_{\tilde{F}}(\boldsymbol{z}) = (\pi(\boldsymbol{z}), \pi \circ F(\boldsymbol{z}))$ is a diffeomorphism (of $\mathbb{R}2n$).

It is not hard to adapt the proof of Proposition 26.1 to the more general:

Proposition 27.2 *There is a homeomorphism between the set of pairs (F, U) where F is a C^1 symplectic twist map of $U \subset T^*M$ and the pairs (S, V), where S is in the set of C^2 real valued functions S on an open set V (diffeomorphic to U) of $M \times M$ satisfying the following:*

(i) The map $\boldsymbol{q} \to \partial_2 S(\boldsymbol{q}, \boldsymbol{Q}_0)$ (resp. $\boldsymbol{Q} \to \partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q})$) is a diffeomorphism of the open set $\{(\boldsymbol{q}, \boldsymbol{Q}_0)\} \cap V$ (resp. $\{(\boldsymbol{q}_0, \boldsymbol{Q})\} \cap V$) of M into $\left(T^*_{\boldsymbol{Q}_0} M\right) \cap U$ (resp. $\left(T^*_{\boldsymbol{q}_0} M\right) \cap U$) for each $\boldsymbol{Q}_0$ (resp. $\boldsymbol{q}_0$.)

(ii) $S(\boldsymbol{q}_0, \boldsymbol{q}_0) = 0$, for a given $\boldsymbol{q}_0$.

This correspondence is given by:

$$F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P}) \Leftrightarrow \begin{cases} \boldsymbol{p} = & -\partial_1 S(\boldsymbol{q}, \boldsymbol{Q})) \\ \boldsymbol{P} = & \partial_2 S(\boldsymbol{q}, \boldsymbol{Q})). \end{cases} \tag{27.2}$$

Remark 27.3 As noted before, if $\tilde{M} \cong \mathbb{R}^n$, we can choose $\tilde{U} = \tilde{M} \times \mathbb{R}^n \cong \mathbb{R}2n$ in the above definition and proposition. In this case Corollaries 26.2 and 26.4 also remain valid.

Exercise 27.4 a) Prove Proposition 27.2. Verify that, although we have written things in local coordinates, everything in Proposition 27.4 has intrinsic meaning (e.g. $\partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q})$*is* an element of $T^*_{\boldsymbol{q}_0} M$, which only depends on the point $\boldsymbol{q}_0$ and not the coordinate system chosen).
b) Prove that if M in Proposition 27.2is the covering space of a manifold N with fundamental group Γ, and if S satisfy $S(\gamma \boldsymbol{q}, \gamma \boldsymbol{Q}) = S(\boldsymbol{q}, \boldsymbol{Q})$ as well as (i) and (ii), then the symplectic twist map that S generates is a lift of a symplectic twist map on N.

Exercise 27.5 Show that the set of C^1 twist maps on a compact neighborhood in the cotangent bundle of a manifold is open (*Hint*: prove first that the twist condition is an open condition).

A. The Standard Map on Hyperbolic Manifolds

The examples of symplectic twist maps in general cotangent bundles will mainly come from the next chapter, as time ε of Hamiltonian system satisfying the Legendre condition. In this section, we generalize the standard map further to cotangents of hyperbolic manifolds. We assume a little background in Riemmannian geometry, some of which we review in 26. Recall that a *hyperbolic manifold* M of dimension n is one that is covered by the hyperbolic half space $\mathbb{H}^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_n > 0\}$ given the Riemmannian metric $ds^2 = \frac{1}{x_n^2} \sum_1^n dx_k^2$ (???), which has constant curvature -1. Geodesics on $\mathbb{H}^n$ are open semi circles or straight lines perpendicular to the boundary $\{x_n = 0\}$. The relevant property of the geometry of $\mathbb{H}^n$, and hence of any hyperbolic manifold, is that the exponential map is a *global* diffeomorphism $\exp : T\mathbb{H}^n \to \mathbb{H}^n \times \mathbb{H}^n$, a corollary of the Hopf-Rinow Theorem (Gallot, Hulin and Lafontaine (1987), Section ???). The generalization of the standard map that we present now is in fact valid for any Riemmanian manifold with this property.

Proposition 27.6 *Let $S : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{R}$ be given by:*

$$S(\boldsymbol{q}, \boldsymbol{Q}) = \frac{1}{2} \mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{Q}) + V(\boldsymbol{q}),$$

where $V : \mathbb{H}^n \to \mathbb{R}$ is some C^2 function, and Dis *is the distance given by the hyperbolic metric. Then S is the generating function for a symplectic twist map that we called the* **generalized standard map** *on $\mathbb{H}^n$. Furthermore, if V is equivariant under a group of isometries Σ of $\mathbb{H}^n$ representing the fundamental group of the hyperbolic manifold $M = \mathbb{H}^n/\Sigma$, then S is the generating function for a lift of a symplectic twist map on T^*M.*

Proof. We show that S complies with the hypothesis of Proposition 27.4. We take $M = \mathbb{H}^n, U = T^*\mathbb{H}^n \cong \mathbb{H}^n \times \mathbb{R}^n$. We now prove that $\boldsymbol{q} \to \partial_2 S(\boldsymbol{q}, \boldsymbol{Q}_0)$ (resp. $\boldsymbol{Q} \to \partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q})$) is a diffeomorphism $\mathbb{H}^n \to \mathbb{R}^n$. In Section 26we remind the reader how the geodesic flow and the exponential map of a Riemmannian manifold can be seen both on the tangent bundle and the cotangent bundle (via the duality given by the Legendre transform). In the cotangent bundle the geodesic flow G^t is the hamiltonian flow with Hamiltonian the dual metric $g(\boldsymbol{q})(\boldsymbol{p}, \boldsymbol{p})$ and the exponential map is $\exp_{\boldsymbol{q}}(\boldsymbol{p}) = \pi \circ G^1(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{Q}(\boldsymbol{q}, \boldsymbol{p})$, where $G^1(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})$. We also prove that, if $(\boldsymbol{q}, \boldsymbol{Q})$ is in the range where $(\boldsymbol{q}, \boldsymbol{p}) \to \boldsymbol{q} \times \exp(\boldsymbol{q}, \boldsymbol{p})$ has an inverse (the case for all $(\boldsymbol{q}, \boldsymbol{Q}) \in \mathbb{H}^n \times \mathbb{H}^n$ here), then:

$$\begin{aligned} \partial_1 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}) &= \frac{-\boldsymbol{p}}{\|\boldsymbol{p}\|} = \frac{-\boldsymbol{p}}{\mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q})} \\ \partial_2 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}) &= \frac{\boldsymbol{P}}{\|\boldsymbol{P}\|} = \frac{\boldsymbol{P}}{\mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q})} \end{aligned} \tag{27.3}$$

and hence $\partial_1 \mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{p}) = -\boldsymbol{p}, \partial_2 \mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{P}$. The assumption that the exponential is a diffeomorphism means, in this notation, that $\boldsymbol{p} \to \boldsymbol{Q}(\boldsymbol{q}_0, \boldsymbol{p})$ is a diffeomorphism for each fixed $\boldsymbol{q}_0$ and G^1 is a symplectic twist map . Likewise $\boldsymbol{P} \to \boldsymbol{q}(\boldsymbol{Q}_0, \boldsymbol{P})$ is a diffeomorphism because G^{-1}, the inverse of asymplectic twist map must be a symplectic twist map itself. Thus we have established that the maps $\boldsymbol{q} \mapsto \partial_2 \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{Q}_0)$ and $\boldsymbol{Q} \mapsto \partial_1 \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}_0, \boldsymbol{Q})$ are both diffeomorphisms for each *fixed* $\boldsymbol{q}_0, \boldsymbol{Q}_0$. Coming back to our generating function, we have proven that:

$$\boldsymbol{q} \mapsto \partial_2 S(\boldsymbol{q}, \boldsymbol{Q}_0) = \partial_2 \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{Q}_0)$$

is a diffeomorphism, and

$$\boldsymbol{Q} \mapsto \partial_1 S(\boldsymbol{q}_0, \boldsymbol{Q}) = \partial_1 \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}_0, \boldsymbol{Q}) + dV(\boldsymbol{q}_0)$$

must also be a diffeomorphism $\mathbb{H}^n \to T_{\boldsymbol{q}_0}\mathbb{H}^n$ since we added a constant translation by $dV(\boldsymbol{q}_0)$ to a diffeomorphism. Proposition 27.4concludes the proof that S is the generating function for a twist map of $T^*\mathbb{H}^n$. The last statement of the proposition is an easy consequence of Exercise 27.4. □

28. Elliptic Fixed Points

As we will see in Appendix 1 or SG, the study of Hamiltonian dynamics around a periodic orbit of a time independent Hamiltonian reduces to that of a symplectic map:

$$\mathcal{R} : \mathbb{R}2n \to \mathbb{R}2n, \quad \text{such that} \quad \mathcal{R}(0) = 0,$$

called the Poincaré return map.

We now follow Moser (1977). If 0 is an elliptic fixed point, that is $D\mathcal{R}(0)$ has all its eigenvalues on the unit circle, a normal form theorem ???(find ref.) says that (generically?) the map $\mathcal{R}$ is, around 0 given by:

$$\begin{aligned} Q_k &= q_k cos\Phi_k(\boldsymbol{q},\boldsymbol{p}) - p_k sin\Phi_k(\boldsymbol{q},\boldsymbol{p}) + f_k(\boldsymbol{q},\boldsymbol{p}) \\ P_k &= q_k sin\Phi_k(\boldsymbol{q},\boldsymbol{p}) + p_k cos\Phi_k(\boldsymbol{q},\boldsymbol{p}) + g_k(\boldsymbol{q},\boldsymbol{p}) \\ \Phi_k(\boldsymbol{q},\boldsymbol{p}) &= \alpha_k + \sum_{l=1}^{n} \beta_{kl}(q_l^2 + p_l^2). \end{aligned}$$

where the error term f_k, g_k are C^3.[5]

We now show how this map is, in "polar coordinates", a symplectic twist map of $T^*\mathbb{T}^n$, *whenever the matrix* $\{\beta_{kl}\}$ *is non singular.*

Let V be a punctured neighborhood of 0 such that: $0 < \sum_k (q_k^2 + p_k^2) < \epsilon$.

We introduce on V new coordinates (r_k, θ_k) by:

$$q_k = \sqrt{2 r_k \epsilon} cos 2\pi\theta_k, \quad p_k = \sqrt{2 r_k \epsilon} sin 2\pi\theta_k$$

where θ_k is determined modulo 1. One can check that V is transformed into the "annular" set:

$$U = \left\{ (\theta_k, r_k) \in Tn \times \mathbb{R}^n \quad \middle| \quad r_k > 0 \quad \text{and} \quad \sum_k r_k < \frac{1}{2} \right\}$$

Since the symplectic form $d\boldsymbol{q} \wedge d\boldsymbol{p}$ is transformed into $2\pi\epsilon d\boldsymbol{r} \wedge d\boldsymbol{\theta}$, $\mathcal{R}$ remains symplectic in these new coordinates, with the symplectic form $d\boldsymbol{r} \wedge d\boldsymbol{\theta}$. In fact, $\mathcal{R}$ is exact symplectic in U. To check this, it is enough to show that, for any closed curve γ:

$$\int_{\mathcal{R}\gamma} \boldsymbol{r} d\boldsymbol{\theta} = \int_{\gamma} \boldsymbol{r} d\boldsymbol{\theta}.$$

(see Exercise 46.7). It is easy to see that $4\pi\epsilon r_k d\theta_k = p_k dq_k - q_k dp_k$, so by Stokes' theorem:

$$4\pi\epsilon \int_{\gamma} \boldsymbol{r} d\boldsymbol{\theta} = \int_{\partial D} \boldsymbol{p} d\boldsymbol{q} - \boldsymbol{q} d\boldsymbol{p} = -2 \int_{D} \omega$$

where D is a 2 manifold in V with boundary $\partial D = \gamma$. Since $\mathcal{R}$ preserves ω in V, it must preserve the last integral, and hence the first.

To see that $\mathcal{R}$ satisfies the two other conditions for being a symplectic twist map, we just write $\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{r}) = (\boldsymbol{\Theta}, \boldsymbol{R})$ in the new coordinates then:

$$\begin{aligned} \Theta_k &= \theta_k + \psi_{F_k}(\boldsymbol{r}) + o_1(\epsilon) \\ R_k &= r_k + o_1(\epsilon) \\ \text{with } \psi_{F_k} &= \alpha_k + \epsilon \sum_{l=1}^{n} 2\beta_{kl} r_l. \end{aligned}$$

where $\epsilon^{-1} o_1(\epsilon, \boldsymbol{\theta}, \boldsymbol{r})$ and its first derivatives in $\boldsymbol{r}, \boldsymbol{\theta}$ tend to 0 uniformally as $\epsilon \to 0$. We can rewrite this as:

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{r}) = (\boldsymbol{\theta} + \epsilon \boldsymbol{B} \boldsymbol{r} + \boldsymbol{\alpha} + o_1(\epsilon), \boldsymbol{r} + o_1(\epsilon)).$$

[5] actually, one only need them to have vanishing derivatives up to order 3 at the origin and be C^1 otherwise.

So for small ϵ, the condition $\det \partial\boldsymbol{\Theta}/\partial\boldsymbol{r} \neq 0$ is given by the nondegeneracy of $\boldsymbol{B} = \{\beta_{kl}\}$, one uses the fact that $\mathcal{R}$ is C^1 close to a completely integrable symplectic twist map to show that $\mathcal{R}$ is twist in U (the twist condition is open.) The fact that it is homotopic to Id derives from Exercise 23.2.

Note that the set V and therefore U are not invariant under $\mathcal{R}$. However, it is still possible to show the existence of infinitely many periodic points for $\mathcal{R}$: this is the content of the Birkhoff– Lewis theorem (???: state it precisely somewhere) (see Moser (1977)) .

Lemma STMdiffeo is 26.3, Exercise STMstmopen is 27.5, example STMstandardexample is 25.1 , Proposition STMsuffstm is 26.2, formerly a Corollary (Coro), Proposition STMlocglobal is 26.4, Section STMsecelliptic is 28.0, Proposition STMpropdiff is 26.5, STMpropactionpr is Proposition 24.1, Exercise STMexohomt is 23.2.

CHAPTER 5 or STMP

PERIODIC ORBITS FOR SYMPLECTIC TWIST MAPS OF T*T^n

September 251999

29. Presentation Of The Results

Rewrite the ghost tori section (too silly!). Points to be made: parallele with Floer homology, possible dynamic use (2 dim case and more). The proof of Corollary 33.2 can be made using Conley theory only: do that if I get rid of Morse theory in TOPO

In this Chapter, we give some results on existence and multiplicity of periodic orbits of different rotation vectors for symplectic twist map of $T^*\mathbb{T}^n$. The introduction of more refined topological tools yield an improvement on the results of Golé (1989)(see also Golé (1991)).

Similarly to the case $n = 1$, a point $(\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}2n$ is called a $\boldsymbol{m}, d$*–periodic point* for the lift F of a map f of $T^*\mathbb{T}^n$ if

$$F^d(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{p})$$

where $\boldsymbol{m} \in \mathbb{Z}^n$ and $d \in \mathbb{Z}^+$. The rational vector $\frac{\boldsymbol{m}}{d}$ is called the *rotation vector* of the orbit of $(\boldsymbol{q}, \boldsymbol{p})$. In general, the rotation vector (when it exists) of a sequence $\{\boldsymbol{q}_k\}_{k\in\mathbb{Z}} \in (\mathbb{R}^n)^{\mathbb{Z}}$ is given by the limit: $\rho(\overline{\boldsymbol{q}}) = \lim_{k\to\pm\infty} \boldsymbol{q}_k$.

The maps that we consider here satisfy either one of the following two assumptions: $F = F_N \circ \ldots \circ F_1$ is the product of lifts of symplectic twist maps of $T^*\mathbb{T}^n$, with generating functions S_k such that either the following convexity or asymptotic linearity conditions:

Convexity There is a positive real a such that:

$$(29.1)\ (a) \qquad \langle \partial_{12} S_k(\boldsymbol{q}, \boldsymbol{Q}).\boldsymbol{v}, \boldsymbol{v}\rangle \leq -a\left\|\boldsymbol{v}\right\|^2, \quad \forall \boldsymbol{q}, \boldsymbol{Q}, \boldsymbol{v} \in \mathbb{R}^n, k \in \{1, \ldots, N\}.$$

Equivalently:

$$(29.1)\ (b) \qquad F_k(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P}) \qquad \text{and} \qquad \left\langle \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\right)^{-1} \boldsymbol{v}, \boldsymbol{v} \right\rangle \geq a\left\|\boldsymbol{v}\right\|^2, \quad \forall \boldsymbol{v} \in \mathbb{R}^n.$$

uniformly in $(\boldsymbol{q}, \boldsymbol{p})$.

Asymptotic Linearity

$$S_k(\boldsymbol{q},\boldsymbol{Q}) = \frac{1}{2}\langle A_k(\boldsymbol{Q}-\boldsymbol{q}),(\boldsymbol{Q}-\boldsymbol{q})\rangle + R_k(\boldsymbol{q},\boldsymbol{Q})$$

with:

(29.2) (a) $$A_k = A_k^t,\ \det A_k \neq 0$$

(29.2) (b) $$\det \sum_1^N A_k^{-1} \neq 0$$

(29.2) (c) $$\lim_{\|\boldsymbol{Q}-\boldsymbol{q}\|\to\infty} \frac{\nabla R_k(\boldsymbol{q},\boldsymbol{Q})}{\|\boldsymbol{Q}-\boldsymbol{q}\|} = 0.$$

Equivalently:

$$F_k(\boldsymbol{q},\boldsymbol{Q}) = (\boldsymbol{q} + A_k^{-1}\boldsymbol{p} + \Theta(\boldsymbol{q},\boldsymbol{p}),\ \boldsymbol{p} + \Upsilon(\boldsymbol{q},\boldsymbol{p}))$$

with (29.2) (a) and (b) holding for A_k and:

(29.2) (c') $$\lim_{\|\boldsymbol{p}\|\to\infty} \frac{\Theta(\boldsymbol{q},\boldsymbol{p})}{\|\boldsymbol{p}\|} = \lim_{\|\boldsymbol{p}\|\to\infty} \frac{\Upsilon(\boldsymbol{q},\boldsymbol{p})}{\|\boldsymbol{p}\|} = 0$$

Theorem 29.1 *Let $F = F_N \circ \ldots \circ F_1$ be a finite composition of symplectic twist maps F_k of $T^*\mathbb{T}^n$ satisfying either the convexity condition (29.1) or the asymptotic condition (29.2) . Then, for each relatively prime $(\boldsymbol{m}, d) \in \mathbb{Z}^n \times \mathbb{Z}$, F has at least $n+1$ periodic orbits of type $\boldsymbol{m}, d$. It has at least 2^n of them when they are all non–degenerate.*

The proof of this theorem appeared in several pieces: the existence in the convex case was given by Kook & Meiss (1989). Their proof of multiplicity was corrected by the author in Golé (1994). The proof of the theorem with the asymptotic condition is the center of the author's thesis Golé (1989)(see also Golé (1991)). The proof we present here is also more unified, and hopefully simpler. It also improves on our previous results where, in certain cases, we could not garantee the existence of more than 2^{n-1} periodic orbits.

Comments on Conditions (29.1) and (29.2) . In Chapter STM, Proposition 26.5, we derived $\frac{\partial Q}{\partial p}(\boldsymbol{q},\boldsymbol{p}) = -\left(\partial_{12}S(\boldsymbol{q},\boldsymbol{Q})\right)^{-1}$, by implicit differentiation of $\boldsymbol{p} = -\partial_1 S(\boldsymbol{q},\boldsymbol{Q})$. The convexity condition (29.1) (a) thus translates to (29.1) (b). Note that (29.1) (b) means that F has bounded, positive definite twist. MacKay & al. (1989) imposed this condition on their definition of symplectic twist maps, a terminology that we have taken from them. Remember that Proposition 26.4in Chapter STM shows that the bounded twist condition (29.2) implies the global twist condition.

As for Condition (29.2) we stress that each A_k is *not necessarily positive definite*, but only a nondegenerate symmetric matrix. This is what Hermann (1990)called the *indefinite case.* If we set $R_k = 0$ in S_k, we obtain a quadratic generating function for a linear symplectic twist map $L_k(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q} + A_k^{-1}\boldsymbol{p},\boldsymbol{p})$. Thus, if $L = L_N \circ \ldots \circ L_1$, condition (29.2) implies that

(29.3) $$L(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q} + A\boldsymbol{p},\boldsymbol{p}) \quad \text{with} \quad A = \sum_{k=1}^{dN} A_k^{-1}$$

is a symplectic twist map. Hence Condition (29.2) can be expressed as saying that F is asymptotically linear (and asymptotically completely integrable), in that it is close to L at ∞: (29.2) (c') shows that

$$\lim_{\|\boldsymbol{p}\|\to\infty} \frac{\|F(\boldsymbol{q},\boldsymbol{p}) - L(\boldsymbol{q},\boldsymbol{p})\|}{\|\boldsymbol{p}\|} = 0.$$

We leave it to the reader to show that the generating function and map conditions in (29.2) are indeed equivalent.

Example 29.2 The generalized standard map satisfies both conditions (29.1) and (29.2)

Outline of the proof. In the convex case, we start by finding a minimum for a discrete action function W, sum of generating functions. The convexity condition, as in the classical calculus of variation gives us coercion on W, which implies the existence of the minimum. The multiplicity is given by Morse theory on an adequately chosen sublevel set $\{W \leq C\}$.

The case with the asymptotic condition is a relatively easy consequence of Proposition 52.8: we find that the action function W on the appropriate quotient space of the space of sequences is indeed quadratic at infinity as required by that Proposition.

30. Finite Dimensional Variational Setting

Let $F = F_N \circ \ldots \circ F_1$ where each F_k is the lift of a symplectic twist map with generating function S_k. The critical action principle in Chapter STM tells us that finding orbits of F can be done by finding solutions of:

$$\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) + \partial_2 S_{k-1}(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k) = 0 \tag{30.1}$$

The appropriate space of sequences in which to look for solutions of (30.1) corresponding to $\boldsymbol{m}, d$–points of F is:

$$\overline{\mathbf{X}} = \{\overline{\boldsymbol{q}} \in (\mathbb{R}^n)^{\mathbb{Z}} \mid \boldsymbol{q}_{k+dN} = \boldsymbol{q}_k + \boldsymbol{m}\}$$

which is isomorphic to $(\mathbb{R}^n)^{dN}$: the terms $(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN})$ determine a whole sequence in $\overline{\mathbf{X}}$, and we will use them as a coordinate system for this space. Finding a sequence satisfying (30.1) in $\overline{\mathbf{X}}$, is equivalent to finding $\overline{\boldsymbol{q}} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN})$ which is a critical point for the function:

$$W(\overline{\boldsymbol{q}}) = \sum_{k=1}^{dN} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}),$$

in which we set $\boldsymbol{q}_{dN+1} = \boldsymbol{q}_1$. In fact, the proof of the critical action principle (see Proposition 24.1and also Corollary 5.2) reduces in this case to the suggestive formula:

$$dW(\overline{\boldsymbol{q}}) = \sum_{k=1}^{dN} (\boldsymbol{P}_{k-1} - \boldsymbol{p}_k) d\boldsymbol{q}_k. \tag{30.2}$$

The search for critical points of W will be made by studying the gradient flow solution of

$$\frac{d\overline{\boldsymbol{q}}(t)}{dt} = -\nabla W(\overline{\boldsymbol{q}}(t))$$

where t is an artificial time variable. Written in components, this equation is the differential equation:

$$\dot{\boldsymbol{q}}_k = -\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) - \partial_2 S_{k-1}(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k)$$

which for C^2 functions S_k's defines a local flow φ^t on $\overline{\mathbf{X}}$. This flow will certainly be defined for all $t \in \mathbb{R}$ whenever the second derivatives of the S_k's are bounded: the vector field $-\nabla W$ is then globally Lipschitz.

We need to complicate matters some more. First, notice that $\overline{\mathbf{X}}$ has trivial topology, so we should take advantage of the periodicity of W. Formally, this can be done by remarking that W is invariant under the diagonal $\mathbb{Z}^n$ action: $W \circ \tau_{\boldsymbol{n}} = W,\ \boldsymbol{n} \in \mathbb{Z}^n$ where

$$\tau_{\boldsymbol{n}}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN}) = (\boldsymbol{q}_1 + \boldsymbol{n}, \ldots, \boldsymbol{q}_{dN} + \boldsymbol{n}).$$

Hence W induces a function on the quotient $\overline{\mathbf{X}}/\mathbb{Z}^n$. This operation takes in account the fact that the maps F and F_k are all lifts of maps of $T^*\mathbb{T}^n$. Without this condition it is easy to find maps of $\mathbb{R}2n$ without $\boldsymbol{m}, d$-orbits,*eg.* $(\boldsymbol{q}, \boldsymbol{p}) \mapsto (\boldsymbol{q}, \boldsymbol{p} + \boldsymbol{a})$.

But we go one step further. We are not satisfied with finding distinct $\boldsymbol{m}, d$–points, but we want to make sure that different critical points of our function W correspond in fact to different $\boldsymbol{m}, d$*–orbits* of F. To this effect, we note that W is also invariant under the N^{th} iterate σ^N of the shift map:

$$(\sigma\overline{\boldsymbol{q}})_k = \boldsymbol{q}_{k+1}.$$

This is because $S_{k+N} = S_k$, and thus σ^N permutes circularly the terms of W. Hence we can define W succesively on the quotients:

$$\begin{aligned} \overline{\mathbf{X}} &= \overline{\mathbf{X}}/\tau = \overline{\mathbf{X}}/\mathbb{Z}^n \quad \text{and} \\ \mathbf{X} &= \overline{\mathbf{X}}/\sigma^N = \overline{\mathbf{X}}/(\mathbb{Z}^n \times \mathbb{Z}) \end{aligned}$$

of $\overline{\mathbf{X}}$ by the actions of $\tau_{\boldsymbol{n}},\ \boldsymbol{n} \in \mathbb{Z}^n$ and σ^N. Since the action of σ^N on critical sequences corresponds to the action of F on points of $T^*\mathbb{T}^n$, distinct critical points of W on $\mathbf{X}$ correspond to distinct orbits of F.

The following lemma, due to Bernstein & Katok (1987), describes the topology of the problem:

Lemma 30.1 *The quotient maps: $\overline{\mathbf{X}} \to \overline{\mathbf{X}}$ and $\overline{\mathbf{X}} \to \mathbf{X}$ are covering maps , and thus so is $\overline{\mathbf{X}} \to \mathbf{X}$. The space $\overline{\mathbf{X}}$ is homeomorphic to $\mathbb{T}^n \times (\mathbb{R}^n)^{dN-1}$, whereas $\mathbf{X}$ is a (not always trivial) fiber bundle with base $\mathbb{T}^n$ and fiber $(\mathbb{R}^n)^{dN-1}$.*

Proof. We make the change of variables:

$$\begin{aligned} \boldsymbol{q} &= \frac{1}{dN}\sum_{1}^{dN} \boldsymbol{q}_k \\ \boldsymbol{v}_k &= \boldsymbol{q}_{k+1} - \boldsymbol{q}_k - \boldsymbol{m}/dN, \quad k \in \{1, \ldots, dN-1\} \end{aligned}$$

and think of $\boldsymbol{q}$ as the base coordinate and $\boldsymbol{v}$ as the fiber. In these coordinates:

$$\begin{aligned} \tau_{\boldsymbol{n}}(\boldsymbol{q}, \boldsymbol{v}) &= (\boldsymbol{q} + \boldsymbol{n}, \boldsymbol{v}) \\ \sigma(\boldsymbol{q}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{dN-1}) &= \left(\boldsymbol{q} + \frac{\boldsymbol{m}}{dN}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{dN-1}, -\sum_{j=1}^{dN-1} \boldsymbol{v}_j\right) \\ \sigma^{dN}(\boldsymbol{q}, \boldsymbol{v}) &= (\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{v}) \end{aligned}$$

(the reader should verify this...) From the first equality, we get:

$$\overline{\mathbf{X}} \stackrel{\text{def}}{=} \overline{\mathbf{X}}/\mathbb{Z}^n \simeq \mathbb{T}^n \times (\mathbb{R}^n)^{dN-1}.$$

and σ^N induces a d–periodic, fixed point free diffeomorphism on $\overline{\mathbf{X}}$, and thus taking the quotient of $\overline{\mathbf{X}}$ by σ^N gives again a covering map. Finally, these coordinates show that $\mathbf{X} = \overline{\mathbf{X}}/\sigma^N$ is a fiber bundle over $(\mathbb{R}^n/\mathbb{Z}^n)/\frac{m}{d}\mathbb{Z} \simeq \mathbb{T}^n$. □

31. Second Variation

In this section, we show how the second derivative of W can be used to decide if a periodic orbit is nondegenerate or not.

Definition 31.1 A periodic point $\boldsymbol{z}$ of period d for a symplectic twist map F is called *nondegenerate* if $DF^d_{\boldsymbol{z}}$ has no eigenvalue 1.

Suppose $F = F_N \circ \ldots \circ F_1$ where each F_k is a symplectic twist map and let W be defined as before.

Lemma 31.2 *An $\boldsymbol{m}, d$ periodic point is nondegenerate for F if and only if the critical point of W to which it corresponds is nondegenerate.*

Proof. Suppose that $(\boldsymbol{q}_1, \boldsymbol{p}_1) = \boldsymbol{z}_1$ is an $\boldsymbol{m}, d$ point for F. We want to solve the equation:

$$DF^d_{\boldsymbol{z}_1}(v) = \lambda v \tag{31.1}$$

with $v \in T(T^*\mathbb{T}^n)_{\boldsymbol{z}_1}$. We follow MacKay & Meiss (1983): If $\overline{\boldsymbol{q}}$ corresponds to the orbit of $\boldsymbol{z}_1$ under the the successive F_k's, it must satisfy:

$$\frac{\partial W(\overline{\boldsymbol{q}})}{\partial \boldsymbol{q}_k} = \partial_2 S_{k-1}(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k) + \partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = 0.$$

Therefore, a "tangent orbit" $\delta\overline{\boldsymbol{q}}$ must satisfy:

$$S^{k-1}_{21}\delta\boldsymbol{q}_{k-1} + (S^k_{11} + S^{k-1}_{22})\delta\boldsymbol{q}_k + S^k_{12}\delta\boldsymbol{q}_{k+1} = 0 \tag{31.2}$$

where we have abbreviated:

$$S^k_{ij} = \partial_{ij} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}).$$

Remark 31.3 This rather physical argument can be given a more mathematical footing. Consider the following:

$$\begin{aligned} T^*\mathbb{R}^n &\cong \left\{((\boldsymbol{q}_1, \boldsymbol{p}_1), \ldots, (\boldsymbol{q}_{dN}, \boldsymbol{p}_{dN}) \in (T^*\mathbb{R}^n)^{dN} \mid F_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{p}_{k+1})\right\} \\ &\cong \{\overline{\boldsymbol{q}} \in (\mathbb{R}^n)^{dN+1} \mid \nabla W(\overline{\boldsymbol{q}})_k = 0, k = 1, \ldots, dN-1\} \end{aligned}$$

The first homeomorphism is between points in the space and their orbit segments of a given length, the second is given by the correspondence between orbit segments and critical points of the action. If one expresses a parametrization of an element of $T(T^*\mathbb{R}^n)$ with the first representation, one gets the orbit of a tangent vector under the differentials of the F_k's. If one uses the second identification , one gets (31.2) .

When $\overline{q}$ corresponds to a periodic point (q_1, p_1), Equation (31.1) translates, in terms of the $\delta\overline{q}$, to:

$$\delta q_{dN+1} = \lambda \delta q_1 \tag{31.3}$$

Equations (31.2) , (31.3) can be put in matrix form as $M(\lambda)\delta\overline{q} = 0$ where $M(\lambda)$ is the following $dNn \times dNn$ matrix:

$$M(\lambda) = \begin{pmatrix} S_{22}^{dN} + S_{11}^1 & S_{12}^1 & 0 & \dots & 0 & \frac{1}{\lambda} S_{21}^{dN} \\ S_{21}^1 & S_{22}^1 + S_{11}^2 & S_{12}^2 & \ddots & & 0 \\ 0 & S_{12}^2 & & & & \vdots \\ \vdots & \ddots & & & & 0 \\ 0 & \dots & 0 & & & S_{12}^{dN-1} \\ \lambda S_{12}^{dN} & 0 & \dots & 0 & S_{21}^{dN-1} & S_{22}^{dN-1} + S_{11}^{dN} \end{pmatrix}$$

(each entries represents an $n \times n$ matrix.) Hence the eigenvalues of $DF^d_{z_1}$ are in one to one correspondence with the values λ for which $\det M(\lambda) = 0$. More precisely, to each eigenvector of $DF^d_{z_1}$ corresponds one and only one vector $\delta\overline{q}$ solution of $M(\lambda)\delta\overline{q} = 0$. Setting $\lambda = 1$, we get $M(1) = \nabla^2 W$, which finishes the proof. □

Remark 31.4 The above relationship between eigenvalues of DF^d and of $\nabla^2 W$ can be given a symplectic interpretation: the Lagrangian manifolds graph(dW) and graph(F) are related by symplectic reduction. Lemma 31.2can then be restated in terms of the invariance of a certain Maslov index under reduction Viterbo (1987).

Lemma 31.2 proves in particular that the condition “all m, d orbits are nondegenerate” is equivalent to “W is a Morse function”. The following proposition shows that both properties are true for generic symplectic twist maps .

Proposition 31.5 *For generic symplectic twist maps , all periodic orbits are nondegenerate and hence all the functions W are Morse*

Proof. We remind the reader that a property is *generic* on a topological space if it satisfied on a *residual* set of that space, *i.e.* a countable intersection of open and dense sets. Robinson Robinson (???), in his theorem 1Bi, proves that the set of C^k symplectic maps with nondegenerate periodic points is residual in the space of all C^k symplectic maps. He proceeds by induction on the period d of the points[(6)]. We want to adapt his proof to the space STM of C^1 of symplectic twist maps . First note that, since the twist condition is open, STM is an open set in the space of C^1 exact symplectic maps. The only thing that we have to check, therefore, is that the perturbations that Robinson uses to kill degeneracy transform exact symplectic maps into exact symplectic maps. But this is not hard to check: each of these perturbations is given by composing the original map f with the time one map of the hamiltonian flow associated to a bump function in a small neibourghood of a given periodic point. Hence the perturbed map is the composition of the original exact symplectic map with the time 1 map of a Hamiltonian, also exact symplectic by Theorem 47.7. The composition of two exact symplectic maps being exact symplectic, we are done. □

---

6 C.Robinson actually deals with higher order resonnances as well, i.e, roots of unity in the spectrum of Df^d_z.

32. The Convex Case

The standing assumption in this section is that $F = F_N \circ \ldots \circ F_1$ where F_k is a symplectic twist map with generating function S_k satisfying the convexity condition:

$$\langle \partial_{12} S_k(\boldsymbol{q}, \boldsymbol{Q}).\boldsymbol{v}, \boldsymbol{v} \rangle \leq -a \|\boldsymbol{v}\|^2, \quad \forall \boldsymbol{q}, \boldsymbol{Q}, \boldsymbol{v} \in \mathbb{R}^n, k \in \{1, \ldots, N\}. \tag{29.1}$$

The central part of the proof of the convex case, due to Kook & Meiss (1989)consists in proving that the function W is proper, and hence has a minimum. This is something we have already done in the case $n = 1$ (see), and the proof in higher dimensions is identical. (??? Change this sentence if I put the min part of AM in a MIN chapter)

Lemma 32.1 *Let S be the generating function of a symplectic twist map satisfying the convexity condition. Then there is an α and positive β and γ such that:*

$$S(\boldsymbol{q}, \boldsymbol{Q}) \geq \alpha - \beta \|\boldsymbol{q} - \boldsymbol{Q}\| + \gamma \|\boldsymbol{q} - \boldsymbol{Q}\|^2 . \tag{32.1}$$

Corollary 32.2 *Let F satisfy the convexity condition (29.2) . Then there is a minimum for the corresponding action function W (and hence an $\boldsymbol{m}, d$–point for F.)*

We have thus found at least one $\boldsymbol{m}, d$–orbit corresponding to a minimum of W. The reader should be aware that, unlike the 1 degree of freedom case, this does not imply that the orbit is a global minimizer (see Hermann (1990) and Arnaud (1989)).

We now turn to the multiplicity of orbits.

This proof can be rewritten using Conley theory only. I should do that if I'm going to get rid of the section on Morse theory in Appendix 2 or TOPO... Outline: Use 51.1(about the retraction): The isolating block W^K with empty exit set, so $H^*(W^K, (W^K)^-) \equiv H^*(W^K)$. Also there is the requisite retraction....

Remember that $\mathbf{X}$ is a bundle over $\mathbb{T}^n$. Let $\Sigma \cong \mathbb{T}^n$ be its zero section. Let $K > \sup_{\overline{\boldsymbol{q}} \in \Sigma} W(\overline{\boldsymbol{q}})$. Trivially, we have:

$$\Sigma \subset W^K \stackrel{\text{def}}{=} \{\overline{\boldsymbol{q}} \in \mathbf{X} \mid W \leq K\}$$

(since W is proper, for almost every K, W^K is a compact manifold with boundary, by Sard's Theorem.) From this we get the commutative diagram in homology:

$$\begin{array}{ccc} H_*(\Sigma) & \xrightarrow{\quad k_* \quad} & H_*(\mathbf{X}) \\ i_* \searrow & & \nearrow j_* \\ & H_*(W^K) & \end{array} \tag{32.2}$$

where i, j, k are all inclusion maps. But $k_* = Id$ since Σ and $\mathbf{X}$ have the same homotopy type. Hence i_* must be injective.

If all the $\boldsymbol{m}, d$–points are nondegenerate, W is a Morse function (a generic situation by Proposition 32.0) and according to Morse Theory (Milnor (1969) , Section 3) W^K has the homotopy type of a finite CW complex, with one cell of dimension k for each critical point of index k in W^K. In particular, we have the following Morse inequalities:

$$\#\{\text{critical points of index } k\} \geq b_k$$

where b_k is the kth Betti number of W^K, $b_k \geq \binom{n}{k}$ in our case since $H_*(\mathbb{T}^n) \hookrightarrow H_*(W^K)$. Hence there are at least 2^n critical points in this nondegenerate case.

If W is not a Morse function, rewrite the diagram (32.2) , but in cohomology, reversing the arrows and raising the stars. Since $k^* = Id$, j^* must be injective this time. We know that the cup length $cl(X) = cl(\mathbb{T}^n) = n+1$. By definition, this means that there are n cohomology classes $\alpha_1, \ldots, \alpha_n$ in $H^1(\mathbf{X})$ such that $\alpha_1 \cup \ldots \cup \alpha_n \neq 0$. Since j^* is injective, $j^*\alpha_1 \cup \ldots \cup j^*\alpha_n \neq 0$ and thus $cl(W^K) \geq n+1$. W^K being compact, and invariant under the gradient flow, Lusternik-Schnirelman theory implies that W has at least $n+1$ critical points in W^K (The proof of Theorem 1 in CH.2 Section19 of Dubrovin & al. (1987) , which is for compact manifolds without boundaries can easily be adapted to this case.) □

33. Asymptotically Linear Systems

In this section we swap the convexity condition (29.1) for asymptotic linearity of the map (29.2) . In this case, the periodic action function W does not necessarily have any minimum. The topological tool we use here is Proposition 52.8.

We remind our reader of our assymption (29.2) : $F = F_N \circ \ldots \circ F_1$ is a product of lifts of symplectic twist maps of $T^*\mathbb{T}^n$. The generating function S_k of F_k satisfies:

$$S_k(\boldsymbol{q}, \boldsymbol{Q}) = \frac{1}{2}\langle A_k(\boldsymbol{Q}-\boldsymbol{q}), (\boldsymbol{Q}-\boldsymbol{q})\rangle + R_k(\boldsymbol{q}, \boldsymbol{Q})$$

with:

$$A_k = A_k^t,\ \det A_k \neq 0, \quad \det \sum_1^N A_k^{-1} \neq 0, \quad \lim_{\|\boldsymbol{Q}-\boldsymbol{q}\|\to\infty} \frac{\nabla R_k(\boldsymbol{q}, \boldsymbol{Q})}{\|\boldsymbol{Q}-\boldsymbol{q}\|} = 0 \tag{29.2}$$

We view R as a *global* perturbation term. As before we let $L_k(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + A_k^{-1}\boldsymbol{p})$ and $L = L_N \circ \ldots \circ L_1$. Then $L(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + A\boldsymbol{p})$ with $A = \sum_1^N A_k^{-1}$. L and all the L_k's are completely integrable symplectic twist maps .

As before, we are looking for critical points of:

$$W(\overline{\boldsymbol{q}}) = \sum_{k=1}^{dN} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = \sum_{k=1}^{dN} \frac{1}{2}\left\langle A_k(\boldsymbol{q}_{k+1}-\boldsymbol{q}_k), (\boldsymbol{q}_{k+1}-\boldsymbol{q}_k)\right\rangle + \sum_{k=1}^{dN} R_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}).$$

where $\overline{\boldsymbol{q}} \in \overline{\mathbf{X}}$ i.e., $\boldsymbol{q}_{dN+1} = \boldsymbol{q}_1$. The first sum in the right hand side is quadratic, call it $\mathcal{Q}'$. It is the action function for the symplectic twist map L defined above. We change coordinates $\Psi : (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN-1}) \mapsto (\boldsymbol{q}, \boldsymbol{v})$ as in Section 30:

$$\boldsymbol{q} = \frac{1}{dN}\sum_1^{dN} \boldsymbol{q}_k$$

$$\boldsymbol{v}_k = \boldsymbol{q}_{k+1} - \boldsymbol{q}_k - \boldsymbol{m}/dN, \quad k \in \{1, \ldots, dN-1\}.$$

In these coordinates, W is of the form:

$$W(\boldsymbol{q}, \boldsymbol{v}) = \mathcal{Q}(\boldsymbol{v}) + R(\boldsymbol{q}, \boldsymbol{v})$$

where $\mathcal{Q}$ is the homogeneous quadratic function:

$$\mathcal{Q}(\boldsymbol{v}) = -\frac{1}{2}\left\langle A_{dN}(\sum_{1}^{dN-1}\boldsymbol{v}_k), \sum_{1}^{dN-1}\boldsymbol{v}_k)\right\rangle + \frac{1}{2}\sum_{k=1}^{dN-1}\langle A_k\boldsymbol{v}_k, \boldsymbol{v}_k\rangle,$$

and $R = \sum_1^{dN} R_k \circ \Psi^{-1}$. Postponing the proof that $\mathcal{Q}(v)$ is nondegenerate, we conclude the proof of the theorem.

The maps τ_n and σ introduced in Section 30 all map fibers to fibers diffeomorphically and linearly in the trivial bundle $\overline{\mathbf{X}} \to \mathbb{R}^n$ with projection $(\boldsymbol{q}, \boldsymbol{v}) \mapsto \boldsymbol{q}$. Hence $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v}) = \mathcal{Q}(\boldsymbol{v})$ which is quadratic nondegenerate in the fibers induces in the quotient $\mathbf{X}$ of $\overline{\mathbf{X}}$ a function $\mathcal{Q}$ which is also quadratic nondegenerate in the fibers of the bundle $\mathbf{X} \to \mathbb{T}^n$. Finally, it is easy to see that the asymptotic condition on R_k given in (29.2) implies that:

$$\frac{1}{\|\boldsymbol{v}\|}\frac{\partial}{\partial \boldsymbol{v}}(W - \mathcal{Q}) = \frac{1}{\|\boldsymbol{v}\|}\frac{\partial R}{\partial \boldsymbol{v}} \to 0 \quad \text{as} \quad \|\boldsymbol{v}\| \to \infty$$

in $\overline{\mathbf{X}}$ and hence also in its quotient $\mathbf{X}$. We apply Proposition gpqi to conclude the proof of Theorem 29.1.

We now turn to the proof that, given the asumption (29.2), $\mathcal{Q}(\boldsymbol{v})$ is nondegenerate. The reader could work the linear algebra out directly. We prefer to give a dynamical argument which might enlight us a bit on the linear asymptotic condition. Critical points of $\boldsymbol{v} \mapsto \mathcal{Q}(\boldsymbol{v})$ form the kernel of $\mathcal{Q}$. On the other hand, critical points of $(\boldsymbol{q}, \boldsymbol{v}) \mapsto \mathcal{Q}(\boldsymbol{q}, \boldsymbol{v}) = \mathcal{Q}(\boldsymbol{v})$ are in one to one correspondence with the $\boldsymbol{m}, d$ orbits of the linear map L. Since L is a linear completely integrable symplectic twist map, these orbits form an n dimensional plane parallele to the 0 section of $T^*\mathbb{T}^n$. Since the generating function of L is quadratic and the above change of coordinate Ψ is affine, this plane corresponds 1-1 to an n-plane of critical points of $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v})$ in $\overline{\mathbf{X}}$. But the n-plane $\{\boldsymbol{v} = 0\}$ is made of critical points of $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v})$. Therefore, there cannot be any other critical points for $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v})$, and hence $\mathcal{Q}(\boldsymbol{v})$ has trivial kernel. □

34. Ghost Tori

Let F be as in Theorem (29.2), and W be the corresponding action function for $\boldsymbol{m}, d$ orbits on $\mathbf{X}$. In the proof of Theorem 29.1(with the asymptotically quadratic condition), we showed that the set of bounded solutions $G = G_1$ of the gradient flow of W continues, in the sense of Conley, the one for the completely integrable map with action function W_0, and that:

$$H^*(G_0) = H^*(\mathbb{T}^n) \hookrightarrow H^*(G)$$

where G_0 is the torus made of critical points of W_0.

Definition 34.1 Let W the action function for a compostion of symplectic twist map $F = F_N \circ \ldots \circ F_1$ on the space $\mathbf{X}$ of $\boldsymbol{m}, d$ sequences. A set G in $\mathbf{X}$ is called a *ghost torus* if it is compact, invariant by the gradient flow of W and if:

$$H^*(\mathbb{T}^n) \hookrightarrow H^*(G).$$

Comments 34.2

(a) If F has an invariant torus made of $\boldsymbol{m}, d$ periodic orbits, the orbit of each point on it corresponds to a critical point in $\mathbf{X}$. Hence the map invariant torus is diffeomorphic to a torus of critical points in $\mathbf{X}$, which is trivially invariant under the gradient flow of W. This torus is hence a ghost torus, we will call it a *completely critical ghost torus* (see Exercise 34.3.)

(b) The spooky connotation in the terminology "ghost tori" can be justified in the following way. One of the essential avenues for the study of symplectic twist maps is the standard family, which fits quite well in the setting of Theorem ???. The paradigm expressed by the standard family is that of a deformation of an integrable map F_0. We have seen that to such a map corresponds a foliation of $T^*\mathbb{T}^n$ by invariant tori, one for each rotation vector. In particular there is exactly one $\boldsymbol{m}, d$ periodic invariant torus for F_0, corresponding to a completely critical ghost torus in the space $\mathbf{X}$ for each $\boldsymbol{m}, d$. One of the fundamental questions in the theory is to understand what happens to these invariant tori as one deforms F_0. ???By now, this should have been stated a hundred times already??? What Theorem ??? shows is that a "ghost" of the invariant torus for F_0 remains, as the parameter s varies, namely G_s, but in the space $\mathbf{X}$. This ghost torus is invariant by the gradient flow of W_s, *but does not necessarily corresponds to an F_s–invariant torus anymore*. Indeed, generically, the only dynamically "visible" part of G_s is formed by the (at least 2^n, but finite number of) critical points that it contains, which correspond to the $\boldsymbol{m}, d$ periodic orbits. G_s is in fact a collection of critical points for W_s and their connecting orbits for the gradient flow : intersections of stable and unstable manifolds for the critical points (this is true of any compact invariant set for a gradient flow.) Here is a table that might be helpful in understanding the analogy we are trying to draw:

Silly Table

Real World	$T^*\mathbb{T}^n, \quad F$
Yonder World	$\mathbf{X} \quad W$
Live Being	Invariant Torus for F
Ghost	Ghost Torus G for $\frac{d}{dt}\overline{\boldsymbol{q}} = \nabla W(\overline{\boldsymbol{q}})$
Soul	$H^*(\mathbb{T}^n) \hookrightarrow H^*(G)$
Time	Parameter in the Standard Map
Transcending	Map $\mathcal{T}$ from $T^*\mathbb{T}^n$ to $\mathbf{X}$:
	$\mathcal{T}(\boldsymbol{q}_1, \boldsymbol{p}_1) = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN})$, where $(\boldsymbol{q}_{k+1}, \boldsymbol{p}_{k+1}) = F_k(\boldsymbol{q}_k, \boldsymbol{p}_k)$.
Appearing	Map $\mathcal{A}$ from $\mathbf{X}$ to $T^*\mathbb{T}^n$:
	$\mathcal{A}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{dN}) = (\boldsymbol{q}_1, \boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2))$.

(c) Instead of thinking of G_s as a subset of $\mathbf{X}$, one can remember that the set G_s is the projection of the τ and σ^N invariant set gG_s in $\overline{\mathbf{X}} \subset (\mathbb{R}^n)^{\mathbb{Z}}$.

(d) If F is as in Theorem 29.1(convex case), one can reword the proof of that theorem in order to deduce the existence of a ghost torus: we have shown in ??? that a map satisfying the convexity condition ??? could be deformed to a completely integrable one, through a path of symplectic twist maps satisfying this condition. Let F_s be such a path and W_s the corresponding action function. Since we have seen in the proof of Theorem ??? that they were no critical points outside of a set W_s^K for K big enough (we can make K uniform in $s \in [0, 1]$), the set G_s of bounded solutions for the gradient flow of W must be included in W_s^K and thus (see ???) the sets G_s are related by continuation. G_0 is normally hyperbolic, as in the proof of Theorem ??? and thus we can conclude this alternate proof of Theorem ??? (convex case) as in ??? (formno...), and in particular $G = G_1$ is a ghost torus for $F = F_1$.

(e) Ghost tori are quite reminiscent of the set of connecting orbits that supports Floer's homology complex , as it is applied to Hamiltonian systems on the cotangent bundle of $\mathbb{T}^n$ (the space that Cieleback (1992)calls X in .) It is quite probable that, at least at the (co)homology level, when the map F is Hamiltonian and satisfies the hypothesis of Theorem ???, these sets are identical.

(f) We put the title "Rational ghost tori" to this section, because they live in spaces of sequences with rotation vector m/d. We will discuss later the occurence of irrational ghost tori ???, and their connection with the KAM and Aubry–Mather theory.

Exercise 34.3 Show that the Transcendence of an F–invariant torus is a completely critical ghost torus. Show that one is a Live Being if and only if one is the Appearence of one's own Transcendence. In general, reread the previous paragraphs and give them more rigorous sense with the help of the maps $\mathcal{A}$ and $\mathcal{T}$.

Theorem thesis is 29.1

Condition STMPtquad is (29.2)

Lemma STMPlemsecvar is 31.2, Proposition 31.5is STMPpropgeneric Condition STMPconv is (29.1) , Lemma STMPlemquadconv is 32.1

CHAPTER 6 or INV (formerly PB)

INVARIANT MANIFOLDS

10/17/99

I just gutted this chapter, formerly PB, of the proof of Poincaré-Birkhoff . I intend to put survey sections on KAM and separatrices and their splitting (first and last sections).Sections will have to be revised.This could include proofs that KAM tori are Lagrangian, and orbits on Lagrangian graphs are minimizers. I might need to move the "ratchet" proposition to Chapter AM

35. The Theory of Kolmogorov–Arnold–Moser

Will contain the precise statement of the theorem, an idea of the scheme of proof, and proofs that these tori are Lagrangian and made of minimizers.

KAM theory, which proves the existence of many invariant tori for systems close to integrable, is one of the greatest achievements in Hamiltonian dynamics. It has historical roots going back to Weierstrass who, in 1878, wrote to S. Kovalevski that he had constructed formal power series for quasi-periodic solutions to the planetary problem. The denominators of the coefficients of these series involved integer combinations of the frequencies of rotation of the planets around the sun, which could be close to zero and hence impeded the convergence of the series. Weierstrass advised Mittag-Leffler to make this problem of convergence a question for a prize sponsored by the king of Sweden. In the 271 pages work (Poincaré (1890)) for which he won the prize, Poincaré does not solve the problem completely, and his tentative answer to the convergence is negative. In Poincaré (1899) , he speculates on the possibility of such a convergence, given appropriate number theoretic conditions, but still deems it improbable. It was therefore a significant event when Arnold (1963) (in the analytic, Hamiltonian context) and Moser (1962) (in the differentiable twist map context) gave, following the ideas of Kolmogorov (1954) a proof of existence of quasi-periodic orbits on invariant tori filling up a set of positive measure in the phase space. We can only give here a very limited account of this complex theory, and refer to Moser (1973) and de la Llave (1993) for introductions as well as Bost (1986) for an excellent survey and bibliography. There are many KAM theorems, the most applicable ones being often the hardest ones to even state. We present here a relatively simple statement, cited in Bost (1986).

Theorem 35.1 (KAM for symplectic twist maps) *Let f_0 be an integrable symplectic twist map of $\mathbb{T}^n \times \mathbb{D}^n$ of the form:*

$$f_0(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}+\omega(\boldsymbol{p}),\boldsymbol{p})$$

where $\mathbb{D}^n$ *is a disk in* $\mathbb{R}^n$ *and* $\omega : \mathbb{D}^n \mapsto \mathbb{R}^n$ *is* C^∞ *(since* f_0 *is twist,* $D\omega$ *is invertible). Let* $\boldsymbol{p}_0$ *be an interior point of* $\mathbb{D}^n$*. Suppose that the following condition is satisfied:*
Diophantine condition: *there are positive constants* τ *and* c *such that:*

$$\forall k \in \mathbb{Z}^{n+1}\backslash\{0\}, \quad \left|\sum_{j=1}^{n} k_j\omega_j(\boldsymbol{p}_0) + k_{n+1}\right| \geq c\left(\sum_{j=1}^{n+1}|k_j|\right)^{-\tau} \tag{35.1}$$

Then there is a neighborhood W *of* f_0 *of* C^∞ *exact symplectic maps such that, for each* $f \in W$*, there exists an embedded invariant torus* $\mathbb{T}_f \simeq \mathbb{T}^n$ *in the interior of* $\mathbb{T}^n \times \mathbb{D}^n$ *such that:*
(i) $\mathbb{T}_f$ *is a* C^∞ *Lagrangian graph over the zero section*
(ii) $f\big|_{\mathbb{T}_f}$ *is* C^∞ *conjugated to the rigid translation by* $\omega(\boldsymbol{p}_0)$
(iii) $\mathbb{T}_f$ *and the conjugacy depend* C^∞ *on* f*.*
Moreover the measure of the complement of the union of the tori $\mathbb{T}_f(\boldsymbol{p}_0)$ *goes to 0 as* $\|f - f_0\|$ *goes to 0.*

Remark 35.2
1) The diophantine condition (35.1) is shared by a large set of vectors in $\mathbb{R}^n$. As an example, when $n = 1$, the set of real numbers $\mu \in [0, 1]$ such that $|\mu - p/q| > K/q^3$ for some K is dense in $[0, 1]$ and has measure going to 1 as K goes to 0.
2) The most common versions of KAM theorems concern Hamiltonian systems with a Legendre condition. In Chapter 6 we show the intimate relationship of such Hamiltonian systems with symplectic twist maps . It therefore comes as no surprise that KAM theorems have equivalents in both categories of systems. Note that there are *isoenergetic* versions of the KAM theorem for Hamiltonian systems, where the existence of many invariant tori is proven in a prescribed energy level.
3) One important contribution in Moser (1962) was his treatment of the finitely differentiable case: he was able to show a version for $n = 1$ (twist maps) where f_0 and its perturbation are $C^l, l \geq 333$ instead of analytic. This was later improved to $l > 3$ and in higher dimension n, to $l > 2n + 1$ (at least if the original f_0 is analytic).
4) There is a version of the KAM for *non symplectic* perturbations of completely integrable maps of the annulus, called the *Theorem of translated curves*, due to Rüssmann (1970). It states that, around an invariant circle for f_0 whose rotation number ω satisfies the diophantine condition (35.1) (only one j in this case), there exists a circle invariant by $t_a \circ f$ for a perturbation f of f_0 and $t_a(x, y) = (x, y + a)$ which has same rotation number as the original (i.e. the map f has flux $-a$).
5) One may wonder if, among all invariant tori of a symplectic twist map close to integrable, the KAM tori are typical. KAM theory says that in measure, they are. However Herman (1992a) (see also Yoccoz (1992)) shows that, for a generic symplectic twist map close to integrable, there is a residual set of invariant tori on which the (unique) invariant measure has a support of Hausdorff dimension 0. Things get even worse when the differential $D\omega$ in Theorem 35.1 is not positive definite: there may be many invariant tori that project onto, but are not graphs over the 0–section, and this for maps arbitrarily close to integrable (see Herman (1992 b)).

6) KAM theory implies the stability of orbits on the KAM tori, hence stability with high probability. But in "real situations" it is impossible to tell whether motion actually takes place on a KAM torus. Nekhoroshev (1977) provides an estimate of how far a trajectory can drift in the momentum direction over long periods of time: If $H(\boldsymbol{q},\boldsymbol{p}) = h(\boldsymbol{p}) + f_\varepsilon(\boldsymbol{q},\boldsymbol{p})$ is a real analytic Hamiltonian function on $T^*\mathbb{T}^n$ with $f_\varepsilon < \varepsilon$ (a small parameter) and $h(\boldsymbol{p})$ satisfies a certain condition (steepness) implied by convexity, then there exist constants ε_0, R_0, T_0 and a such that, if $\varepsilon < \varepsilon_0$, one has:

$$|t| \leq T_0 \exp[(\varepsilon_0/\varepsilon)^a] \Rightarrow |\boldsymbol{p}(t) - \boldsymbol{p}(0)| \leq R_0(\varepsilon/\varepsilon_0)^a.$$

With a (quasi) convexity condition instead of the steepness condition, Lochak (1992) and Pöshel (1993) showed that the optimal a is $\frac{1}{2n}$. Delshams & Gutiérrez (1996a)presents unified proofs of the KAM theorem and Nekhoroshev estimates for analytic Hamiltonians.

Whereas we cannot give a proof of the KAM theorem in this book, the following theorem (Arnold (1983)) offers a simple model in a related situation in which the KAM method can be applied in a less technical way.This will allow us to sketch very roughly the central ideas of the method.

Theorem 35.3 *There exists $\varepsilon > 0$ depending only on K, ρ and σ such that, if a is a 2π –periodic analytic function on a strip of width ρ, real on the real axis with $a(z) < \varepsilon$ on the strip and such that the circle map defined by*

$$f : x \mapsto x + 2\pi\mu + a(x)$$

is a diffeomorphism with rotation number μ satisfying the diophantine condition:

$$|\mu - p/q| > \frac{K}{q^{2+\sigma}}, \quad \forall\ p/q \in \mathbb{Q}$$

then f is analytically conjugate to a rotation R of angle $2\pi\mu$

Sketch of proof: We seek a change of coordinates $H : \mathbb{S}^1 \to \mathbb{S}^1$ such that:

$$H \circ R = f \circ H \tag{35.2}$$

write $H(z) = z + h(z)$, with $h(z + 2\pi) = h(z)$. Then (35.2) is equivalent to

$$h(z + 2\pi\mu) - h(z) = a(z + h(z)). \tag{35.3}$$

Since $a(z) < \varepsilon$, h must be of order ε as well and thus, in first approximation, (35.3) is equivalent to:

$$h(z + 2\pi\mu) - h(z) = a(z) \tag{35.4}$$

Decomposing $a(z) = \sum a_k e^{i2\pi kz}$, $h(z) = \sum b_k e^{ikz}$ in their Fourier series and equating coefficients on both sides of (35.4) we obtain:

$$b_k = \frac{a_k}{e^{i2\pi k\mu} - 1}$$

where we see the problem of small divisors arise: the coefficients b_k of h may become very big if μ is not sufficiently rational.

It turns out that, assuming the diophantine condition and using an infinite sequence of approximate conjugacies given by solutions of (35.4) , one obtains sequences h_n, a_n and corresponding $H_n, f_n = H_n^{-1} \circ f \circ H_n$ which converge to H, R for some H. The domain of h_n and f_n is a strip that shrinks with n but in a controllable way. This iterative process of "linear" approximations to the conjugacy can be interpreted as a type of Newton's method for the implicit equation $\mathcal{F}(f, H) = H^{-1} \circ f \circ H = R$ (given f, find H) and inherits the quadratic convergence of the classical Newton's method: $R - \mathcal{F}(f_n, H_n) = O(\varepsilon^{2n})$ (see Hasselblat & Katok (1995) Section 2.7.b). □

36. Properties of Invariant Tori

The previous section showed the existence of many invariant tori for symplectic twist maps close to integrable. These tori are Lagrangian graphs with dynamics conjugated to quasi–periodic translation. In dimension 2, the Aubry-Mather theorem gives an answer to the question of what happens to these tori when they break down, in large perturbation of integrable maps. In higher dimension, Mather's theory of minimal measure also provides an answer to that question (see Chapter AMG). In this section, we look for properties that invariant tori may have whether they arise from KAM or not. We will see that certain attributes of KAM tori (*eg.* graphs with recurrent dynamics) imply their other attributes (*eg.* Lagrangian), as well as other properties not usually stated by the KAM theorems (minimality of orbits).

Recurrent Invariant Toric Graphs are Lagrangian

Theorem 36.1 (Hermann (1990)) *Let T be an invariant torus for a symplectic twist map f of $T^*\mathbb{T}^n$ and suppose $f\big|_T$ is conjugated by a diffeomorphism h to a an irrational translation R on $\mathbb{T}^n$. Then T is Lagrangian.*

Proof. Since the 2-form $\omega\big|_T$ is invariant under $f\big|_T$ and since $R = h^{-1} \circ f\big|_T \circ h$, the 2-form $h^*\omega\big|_T$ is invariant under R. Since R is recurrent, $h^*\omega\big|_T = \sum_{i,j} a_k j dx_k \wedge dx_j$ must have constant coefficients a_k. Integrating $h^*\omega\big|_T$ over the x_k,x_j subtorus yields on one hand a_{kj}, on the other hand 0 by Stokes' theorem since $h^*\omega\big|_T = dh^*\lambda\big|_T$ is exact. Hence $h^*\omega\big|_T = 0 = \omega\big|_T$ and the torus T is Lagrangian. □

Orbits on Lagrangian Invariant Tori as Minimizers

The following theorem is attributed to Herman by MacKay & al. (1989), whose proof we reproduce here.

Theorem 36.2 *Let T be torus, C^1 graph over the zero section of $T^*\mathbb{T}^n$ which is invariant for a symplectic twist map f which satisfies the convexity condition (29.1) :*

$$\langle \partial_{12} S_k(\boldsymbol{q}, \boldsymbol{Q}).\boldsymbol{v}, \boldsymbol{v} \rangle \leq -a \left\| \boldsymbol{v} \right\|^2, \quad \forall \boldsymbol{q}, \boldsymbol{Q}, \boldsymbol{v} \in \mathbb{R}^n, k \in \{1, \ldots, N\}.$$

Then any orbit on T is minimizing.

Proof. Since T is Lagrangian, it is the graph of the differential of some function plus a constant 1-form: $T = dg(\mathbb{T}^n) + \beta$ (see SGexolagraph). Let $\psi(\boldsymbol{q}) = \pi f(\boldsymbol{q}, dg(\boldsymbol{q}) + \beta)$ and

$$R(\boldsymbol{q},\boldsymbol{Q}) = S(\boldsymbol{q},\boldsymbol{Q}) + g(\boldsymbol{q}) - g(\boldsymbol{Q}) + \beta(\boldsymbol{q}-\boldsymbol{Q}).$$

We now show that R is constant on T, where it attains its minimum. Following Mather, we first note that:

$$\partial_1 R(\boldsymbol{q},\boldsymbol{Q}) = \partial_1 S(\boldsymbol{q},\boldsymbol{Q}) + dg(\boldsymbol{q}) + \beta = 0 \Leftrightarrow \boldsymbol{p} = dg(\boldsymbol{q}) + \beta \Leftrightarrow \boldsymbol{Q} = \psi(\boldsymbol{q})$$
$$\partial_2 R(\boldsymbol{q},\boldsymbol{Q}) = \partial_2 S(\boldsymbol{q},\boldsymbol{Q}) - dg(\boldsymbol{Q}) - \beta = 0 \Leftrightarrow \boldsymbol{P} = dg(\boldsymbol{Q}) - \beta \Leftrightarrow \boldsymbol{Q} = \psi(\boldsymbol{q})$$

Hence $R(\boldsymbol{q},\psi(\boldsymbol{q})) = R_0$ is constant , and $DR(\boldsymbol{q},\boldsymbol{Q})$ is non zero if $\boldsymbol{Q} \neq \psi(\boldsymbol{q})$. In Lemma STMPlemconvquad, we proved that, when (29.1) holds, the generating function satisfies the following quadratic growth:

$$S(\boldsymbol{q},\boldsymbol{Q}) \geq \alpha - \beta \|\boldsymbol{q}-\boldsymbol{Q}\| + \gamma \|\boldsymbol{q}-\boldsymbol{Q}\|^2 . \tag{36.1}$$

where γ is given by $a/2$ in the convexity condition. Since $\partial_{12}R = \partial_{12}S$, the same kind of quadratic estimate holds for R which is thus bounded below. Since R has all its critical points on T, it must attain its minimum there. It is now easy to see that the $\boldsymbol{q}$ coordinates $\boldsymbol{q}_n, \ldots, \boldsymbol{q}_k$ of any orbit segment on T must minimize the action. Let $r_n, \ldots, r_k$ another sequence of points of $\mathbb{T}^n$ with $\boldsymbol{q}_n = r_n$, $\boldsymbol{q}_k = r_k$. Then:

$$\begin{aligned} W(r_1,\ldots,r_k) &= \sum_{j=n}^{k-1} R(r_j, r_{j+1}) + g(\boldsymbol{q}_k) - g(\boldsymbol{q}_n) + \beta(\boldsymbol{q}_k - \boldsymbol{q}_n) \\ &\geq (k-n)H_0 + g(q_k) - g(q_n) + \beta(\boldsymbol{q}_k - \boldsymbol{q}_n) = W(q_1,\ldots,q_k) \end{aligned}$$

□

Remark 36.3 Arnaud (1989) (see also Hermann (1990)) has interesting examples which show that the condition that the graph be Lagrangian is essential in Theorem 36.2. Consider the Hamiltonians on $T^*\mathbb{T}^2$ is given by:

$$H_\varepsilon(q_1,q_2,p_1,p_2) = \frac{1}{2}(p_1 - \varepsilon\cos(2\pi q_2))^2 + \frac{1}{2}p_2^2.$$

The torus $\{(q_1, q_2, \varepsilon\cos(2\pi q_2), 0)\}$ is made of fixed points for the corresponding Hamiltonian system, but it is not Lagrangian (exercise). A further perturbation $G_{\varepsilon,\delta}(\boldsymbol{q},\boldsymbol{p}) = H_\varepsilon(\boldsymbol{q},\boldsymbol{p}) + \delta\sin(2\pi q_2), 0 < \delta \leq \varepsilon$ of these Hamiltonians also provide counterexamples to the strict generalization of the Aubry-Mather theorem to higher dimensions: such systems have no minimizers of rotation vector 0. All the fixed points for the time 1 map have non trivial elliptic part.

Graph Theorem

Theorem 36.4 (Birkhoff) *Let f be a twist map of the cylinder $\mathcal{A}$. Then:*

(1) **(Graph Theorem)** *Any invariant circle which is homotopic to the circle $C_0 = \{y = 0\}$ is a (Lipschitz) graph over C_0.*

(2) If two invariant circles C_- and C_+ homotopic to $\{y = 0\}$ bound a region without other invariant circles, for any ϵ, there are (uncountably many) orbits going from ϵ-close to $C_\pm$ to ϵ-close to $C_\mp$.

This theorem was proved as two independent theorems by Birkhoff (1920).

Proof. For both (1) and (2), we can assume the existence of an invariant circle, say C_+. Take any circle C which is a graph over C_0 and which lies under C_+ . The image $f(C)$ of this circle may not be a graph anymore, but one can make a *pseudo-graph* $UF(C)$ by *trimming* it: take all the points of $F(C)$ that can be "seen" vertically from above. This set forms the graph of a function which is continuous except for at most countably many jump discontinuities. Because of the positive twist condition, these jumps must always be downward as x increases: if C is given the rightward orientation, a vector tangent to C must avoid the cone Θ_v^+, by the ratchet phenomenon. Make a circle out of this graph by adjoining vertical segments at the jumps. This is $UF(C)$. We call such a curve a *right pseudograph*: a curve made of the graph of a function $y = h(x)$ which is continuous except for downward jump discontinuities (the limit to the right $h(x^+)$ and the left $h(x^-)$ exist at each point and $h(x^-) \geq h(x^+)$), and by adjoining to this graph vertical segments to close the jumps.

We can apply F to a pseudograph C and trim it as we did for a graph. Because of the positive twist condition, the horizontal part of $UF(C)$ is made of images under F of horizontal parts of C. Given a (right pseudo) graph C, we obtain a sequence of curves $C_n = (UF)^n C$.

Lemma 36.5 *$C_\infty = \limsup C_n$ is an f-invariant graph, where $\limsup$ is taken in the sense of functions $y = h(x)$ with the obvious allowance for vertical segments.*

Proof. After one iteration of $U \circ F$ on a (right pseudo) graph C, we get a pseudograph with a *downward modulus of continuity*: the ratchet phenomenon and the vertical cuts implies that, for any pair of points z and z' in the lift of $U \circ F(C)$, $z' - z$ is in a cone V of vectors (x, y) with $y \geq \delta x$ if $x \leq 0$ and $y \leq \delta x$ if $x > 0$ (see Figure 36. 1). This implies that C_∞ also has this modulus of continuity, and hence is a pseudograph.

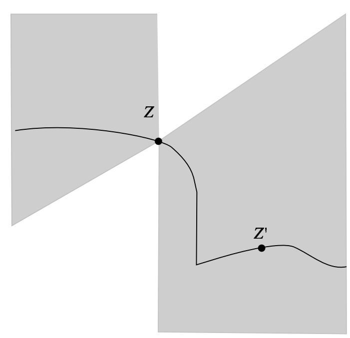


Fig. 36. 1. The cone defining the modulus of continuity at a point z of $U \circ F(C)$.

There is a partial order on circles homotopic to $\{y = 0\}$: we say that $C \preceq C'$ if C' is in the closure of the upper component of $\mathcal{A} \backslash C$, which we denote by $\mathcal{A}_+(C)$. Clearly F and U preserve this order, and $C \preceq U(C)$ for any circle C homotopic to $\{y = 0\}$. This implies that $F^n(C) \preceq UF^n(C) \preceq C_\infty$ for all n, and hence $F(C_\infty) \preceq UF(C_\infty) \preceq C_\infty$. By area preservation $F(C_\infty) = UF(C_\infty) = C_\infty$.

If C_∞ were not a graph, its vertical segments would be mapped by F inside $\mathcal{A}_-(C_\infty) = \mathcal{A}_-(UF(C_\infty))$, and $\mathcal{A}_-(C_\infty)$ would contain $\mathcal{A}_-(F(C_\infty))$ as a proper subset. This contradicts the fact that f has zero flux. Hence C_∞ is an f-invariant graph. □

We now finish the proof of Birkhoff's theorems. Suppose that f admits an invariant circle C_0 homotopic to the boundaries. We show that it is a (Lipschitz) graph. The region below C_0 is invariant. Let C_{max} be the supremum of the invariant *graphs* in this region (under the partial order $\prec$). By continuity, C_{max} is an invariant circle and by Proposition 12.3, it is a Lipschitz graph. If $C_{max} \neq C_0$, then there exist a (not invariant) graph C with $C_{max} \prec C \prec C_0$. Applying the trimming iteration process to C, we get an invariant (Lipschtitz) graph C_∞ with $C_{max} \prec C_\infty \prec C_0$. This contradicts the maximality of C_{max}. Hence $C_0 = C_{max}$ is a Lipschitz graph.

If f does not admit any other invariant circle homotopic to the boundaries than the boundaries themselves, the iteration process performed on any (right) pseudograph must converge to the upper boundary: we have $C \prec UF(C)$. Since $C_\infty \subset closure(\cup f^n(C_0))$, on any graph ϵ close to the lower boundary, there is a point whose ω-limit set is in the upper boundary. We could have defined a trimming L of curves homotopic to the boundaries by taking their lower envelope (the points seen from below) instead of U. Then $L(C)$ is a *left pseudograph* and L preserves the order of circles and $L(C) \prec C$ for any curve C homotopic to the boundaries. Using $\liminf$ instead of $\limsup$ in the argument above, we get an iteration process $L \circ F$ which converges to an invariant graph, which must be the lower boundary this time. And on any graph ϵ close to the upper boundary, there is a point whose ω-limit set is in the lower boundary. □

Remark 36.6 Performing both the $U \circ F$ and $L \circ F$ trimming processes on the same curve C yields points that come arbitrarily close to both boundaries in *forward time*. This fact was proven by Mather (1993) variationally and Hall (1989) topologically. See also LeCalvez (1990) . The results of Mather and Hall are actually sharper as they find orbits whose α-limit set is in one boundary, the ω- limit set in the same or the other boundary (??? check this!). Moreover they find orbits "shadowing" any prescribed sequence of Aubry-Mather sets in a region of instability. It would be interesting to find a new proof of these results based on the trimming technique used above. It would be interesting to generalize the trimming process to Lagrangian pseudographs in higher dimensions.

*. AUBRY-MATHER THEOREM VIA TRIMMING

The above proof of Birkhoff's theorems appears as an aside in Katznelson-Ornstein's paper. They also recover the Aubry-Mather theorem with their trimming method. For this they define, abstractly, a new type of trimming operator, that they call *proper trimming*: one which is such that the area below a curve is preserved under trimming. The main difficulty is to show the existence of such an operator. Once the existence is established, one takes limits of iterations under the map and the trimming operator. The limit is a pseudograph whose horizontal parts are *forward* invariant under f. The Aubry-Mather sets are the intersection of all the forward images of these horizontal parts. Finally, they show the existence of Aubry-Mather sets of all rotation numbers by applying this trimming procedure simultaneously to all the horizontal circles in the annulus. Fathi (???am)offers some relatively distant analog to this in higher dimension, by considering a certain flow on graphs of differentials on cotangent bundles, and recovering the generalized Aubry-Mather sets in the limit.

*. Generalizations of Birkhoff's Graph Theorem to Higher Dimensions

This section surveys (in an all too brief manner) the important work of Bialy, Polterovitch and, indirectly Herman, on invariant Lagrangian tori. It will require from the reader knowledge of material dispersed throughout the book, and more. Bialy & Polterovitch (1992) prove the following generalization to Birkhoff's Graph Theorem. We explain the terminology in the sequel.

Theorem 36.7 *Let F be the time one map of an optical Hamiltonian system of $T^*\mathbb{T}^n$, and let L be a smooth invariant Lagrangian torus for F which satisfies the following conditions:*
1) L is homologous to the zero section of $T^*\mathbb{T}^n$.
2) $F\big|_L$ is either chain recurrent or preserves a measure which is positive on open sets. Then L is a smooth graph (i.e. a section) over the 0-section.

Optical (see Chapter 6) means that the Hamiltonian H is time periodic and convex in the fiber: H_{pp} is positive definite. Homologous to the zero section means that both the invariant torus and the 0-section, seen as homology cycles (which they are because they have empty boundaries) bound a chain of degree $n+1$, presumably some smooth manifold of dimension $n+1$ in our case. As for Condition 2), it suffices here to say that either chain recurrence or existence of an invariant Borel measures are satisfied when the invariant torus is of the type exhibited by the KAM theorem, where the map $F\big|_L$ is conjugated to an irrational translation. In their paper, the authors use a more general condition than 2), which we show at the end of this section is implied by it: *2') the suspension of $F\big|_L$ admits no transversal codimension 1 cocycle homologous to zero.*

This theorem is a culmination of efforts by these authors, as well as Hermann (1990) who gives a perturbative version of this result as some important *a priori* Lipschitz estimates for invariant Lagrangian tori. We now give a very rough idea of the proof of Theorem 36.7. First reduce the theorem to the case of an autonomous Hamiltonian on $T\mathbb{T}^{n+1}$ by viewing time as an extra 1 dimension, with the energy as its conjugate momentum (extended phase space). Assume by contradiction that the invariant torus L is *not* a graph. Consider the set $S(L)$ of critical points of the projection $\pi\big|_L$. Generically, $S(L)$ consists of an $n-1$ dimension submanifold of L whose boundary is of dimension no more than $n-3$. Assume we are in the generic case. Then $S(L)$ can be cooriented by the flow: the Hamiltonian vector field is transverse to it. This makes $S(L)$ a cocycle, i.e. a representent of a cohomology class. It turns out that this cohomology class is dual to the Maslov class of the torus L. The *Maslov class* of L is the pull-back of the generator of $H_1(\Lambda(n))$ by the Gauss map, where $\Lambda(n)$ is the (Grassmanian) space of all Lagrangian planes in $\mathbb{R}2n$. Prosaically, this means the following: *the oriented intersection of $S(L)$ with any closed curve on L counts how many "turns" the Lagrangian tangent spaces of L makes along the curve*. We explain that a little. The number of turns can be made quite precise because $\Lambda(n)$ has one "hole" around which Lagrangian spaces can turn ($H_1(\Lambda(n)) = \mathbb{Z}$). $S(L)$ is the set of points on L where the Lagrangian tangent space becomes vertical in some direction. The tangent space, seen as a graph over the vertical fiber, is given by a bilinear form which is degenerate at points of $S(L)$ and, thanks to the optical condition, decreases index (*i.e.* the dimension of the positive definite subspace increases) when following the flow at those points.

The authors refer to Viterbo (1989) who proves that tori homologous to the zero section have Maslov class zero. Condition 2') now concludes: since it is homologous to zero, the cocycle $S(L)$ must be empty, $i.e.$ there are no singularity in the projection $\pi\big|_L$ and the torus is a graph. The non generic case follows by making a limit argument using uniform Lipschitz estimates for invariant tori proven by Hermann (1990).
Finally, let us show how the fact that $F\big|_L$ is measure preserving implies Condition 2'). Assume F is the time 1 map of an autonomous Hamiltonian system on $T^*\mathbb{T}^n$, L is an invariant torus and Ω is the volume form on L preserved by the Hamiltonian vector field X_H. The Homotopy Formula SGformhomotopy $L_{X_H}\Omega = di_{X_H}\Omega + i_{X_H}d\Omega$ implies that $di_{X_H}\Omega = 0$. Assume X_H is transversal to S, a codimension 1 cocycle homologous to zero and let C be an n-dimensional chain that S bounds. Transversality implies $\int_S i_{X_H}\Omega \neq 0$. On the other hand, Stokes' Theorem yields $\int_S i_{X_H}\Omega = \int_C di_{X_H}\Omega = 0$. This contradiction implies that $S = \emptyset$. □

Remark 36.8 As noted by the authors, it is not clear that Theorem 36.7 is optimal: Condition 2) maybe unnecessary, as is the case in dimension 2. One could imagine a new proof of this theorem using higher dimensional trimming on Lagrangian pseudographs, which would not need this hypothesis...

37. (Un)Stable Manifolds and Heteroclinic orbits

(Un)stable Manifolds

Consider two hyperbolic fixed point $\boldsymbol{z}^* = (\boldsymbol{q}^*, \boldsymbol{p}^*)$, $\boldsymbol{z}^{**} = (\boldsymbol{q}^{**}, \boldsymbol{p}^{**})$ for a symplectic twist map F of $T^*\mathbb{T}^n$. We remind the reader that the *stable and unstable manifolds* at any fixed point $\boldsymbol{z}^*$ are defined as:

$$\mathcal{W}^s(\boldsymbol{z}^*) = \{\boldsymbol{z} \in T^*\mathbb{T}^n \mid F^n(\boldsymbol{z}) = \boldsymbol{z}^*\}, \qquad \mathcal{W}^u(\boldsymbol{z}^*) = \{\boldsymbol{z} \in T^*\mathbb{T}^n \mid F^{-n}(\boldsymbol{z}) = \boldsymbol{z}^*\}$$

Moreover the tangent space to $\mathcal{W}^s$ at $\boldsymbol{z}^*$ is given by the vector subspace $E^s(\boldsymbol{z}^*)$ of eigenvectors of eigenvalue of modulus less than 1, with a similar fact for $\mathcal{W}^u$ and E^u. In our case, the differential DF at the points $\boldsymbol{z}^*$ and $\boldsymbol{z}^{**}$ has as many eigenvalues of modulus less than 1 as it has of modulus greater than 1. Hence the stable and unstable manifolds at these points have both dimension n. The following appears in Tabacman (1993):

Proposition 37.1 *The (un)stable manifolds of a hyperbolic fixed point for a symplectic twist map are Lagrangian. Close to the hyperbolic fixed point, they are graphs of the differentials of functions.*

Proof. Consider a point $\boldsymbol{z}$ on the stable manifold of the hyperbolic fixed point $\boldsymbol{z}^*$, and two vectors $\boldsymbol{v}, \boldsymbol{w}$ tangent to that manifold at $\boldsymbol{z}$. Then:

$$\omega_{\boldsymbol{z}}(\boldsymbol{v}, \boldsymbol{w}) = \omega_{F^k(\boldsymbol{z})}(DF^k(\boldsymbol{v}), DF^k(\boldsymbol{w})) \to \omega^*_{\boldsymbol{z}}(0,0) = 0, \text{as } k \to \infty.$$

which, since it has dimension n in $T^*\mathbb{T}^n$, proves that the stable manifold is Lagrangian. The same argument, using F^{-k}, applies to show that the unstable manifold is Lagrangian. We leave the proof of the second statement to the reader (Exercise 37.2). □

In Exercise INVexoexactstabw, the reader will show a generalization of this fact that makes it applicable to exact symplectic maps (not necessarily twist) of general cotangent bundles.

Variational Approach to Heteroclinic Orbits

As a consequence of Proposition 37.1, we obtain a variational approach to heteroclinic orbits. Let $\boldsymbol{z}^* = (\boldsymbol{q}^*, \boldsymbol{p}^*)$ be a hyperbolic fixed point. Let Φ^u, Φ^s defined on a neighborhood U of $\boldsymbol{q}^*$ be the functions whose differentials define the (un)stable manifolds of $\boldsymbol{z}^*$. We can add appropriate constants to these functions and get $\Phi^s(\boldsymbol{q}^*) = \Phi^u(\boldsymbol{q}^*) = 0$. In the proof of Theorem 36.2, we showed that the function $R(\boldsymbol{q}, \boldsymbol{Q}) = S(\boldsymbol{q}, \boldsymbol{Q}) + g(\boldsymbol{q}) - g(\boldsymbol{Q}) + \beta(\boldsymbol{q} - \boldsymbol{Q})$ was constant on the Lagrangian manifold $Graph(dg + \beta)$. Applying this to $g = \Phi^s$ or $\Phi^u, \beta = 0$, we obtain

$$S(\boldsymbol{q}, \boldsymbol{Q}) = \Phi^s(\boldsymbol{Q}) - \Phi^s(\boldsymbol{q}) + \text{constant},$$

where $F(\boldsymbol{q}, \Phi^s(\boldsymbol{q})) = (\boldsymbol{Q}, \Phi^s(\boldsymbol{Q}))$ (this makes sense in a subset of U). Applying the equation to $(\boldsymbol{q}^*, \boldsymbol{q}^*)$ shows that the constant is $S(\boldsymbol{q}^*, \boldsymbol{q}^*)$. Hence

$$S(\boldsymbol{q}, \boldsymbol{Q}) - S(\boldsymbol{q}^*, \boldsymbol{q}^*) = \Phi^s(\boldsymbol{Q}) - \Phi^s(\boldsymbol{q})$$

for a point $(\boldsymbol{q}, \boldsymbol{Q})$ on the local stable manifold of $\boldsymbol{z}^*$. We now sum over the orbit $(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ of the point $(\boldsymbol{q}, \boldsymbol{Q}) = (\boldsymbol{q}_0, \boldsymbol{q}_1)$ to get:

$$\sum_{k=0}^{N-1} [S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) - S(\boldsymbol{q}^*, \boldsymbol{q}^*)] = \sum_{k=0}^{N-1} [\Phi^s(\boldsymbol{q}_{k+1}) - \Phi^s(\boldsymbol{q}_k)] = \Phi^s(\boldsymbol{q}_N) - \Phi^s(\boldsymbol{q}_0)$$

As $N \to \infty, \Phi^s(\boldsymbol{q}_N) \to \Phi(\boldsymbol{q}^*) = 0$ and thus the sum converges to $-\Phi(\boldsymbol{q}_0)$:

$$\sum_{k=0}^{\infty} [S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) - S(\boldsymbol{q}^*, \boldsymbol{q}^*)] = -\Phi^s(\boldsymbol{q}_0). \tag{37.1}$$

Applying the same manipulations to the unstable manifold, using the fact that the generating function for F^{-1} is $-S(\boldsymbol{Q}, \boldsymbol{q})$, this leads to

Proposition 37.2 *Let $\boldsymbol{z}^* = (\boldsymbol{q}^*, \boldsymbol{p}^*), \boldsymbol{z}^{**} = (\boldsymbol{q}^{**}, \boldsymbol{p}^{**})$ be two hyperbolic fixed points for the symplectic twist map F. Let U^* and U^{**} be neighborhoods of $\boldsymbol{q}^*$ and $\boldsymbol{q}^{**}$ on which the differentials of the functions Φ^u and Φ^s respectively give the unstable manifold of $\boldsymbol{z}^*$ and the stable manifold of $\boldsymbol{z}^{**}$. Then critical points of the function*

$$W(\boldsymbol{q}_0, \ldots, \boldsymbol{q}_N) = \Phi(\boldsymbol{q}_0) + \sum_{k=0}^{N-1} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) - \Phi^s(\boldsymbol{q}_N), \qquad \boldsymbol{q}_0 \in U^*, \boldsymbol{q}_N \in U^{**}$$

are segments of heteroclinic orbits.

Proof. Left to the reader.

With this set-up, Tabacman (1995) shows that, in the 2 dimensional case, any two local minima (*i.e.* fixed points) ξ and η of $\phi(x) = S(x, x)$ such that $\phi(\xi) = \phi(\eta) < \phi(x)$ for all $x \in (\xi, \eta)$, are joined by some trajectory.

Here is a sketch of a numerical algorithm also proposed (and used) by E. Tabacman to find heteroclinic orbits between two given hyperbolic fixed points $\boldsymbol{z}^*, \boldsymbol{z}^{**}$:

(1) Find a basis for the unstable plane E^u of DF at $\boldsymbol{z}^*$, and display the basis vectors as columns of a $2n \times n$ matrix $\begin{pmatrix} A \\ B \end{pmatrix}$

(2) The matrix $M = BA^{-1}$ is symmetric and E^u is the graph of the differential of the quadratic form $\boldsymbol{q} \mapsto \boldsymbol{q}^t M \boldsymbol{q}$. This function is an approximation to Φ^u (see SGexolagsym.)

(3) Perform similar steps to approximate Φ^s at $\boldsymbol{z}^{**}$.

(4) Pick N (large enough) and use your favorite numerical method to search for critical points of the function W defined above, with points $\boldsymbol{q}_0, \boldsymbol{q}_N$ suitably close to $\boldsymbol{z}^*$ and $\boldsymbol{z}^{**}$ respectively.

(5) For more precision, make $\boldsymbol{q}_0$ and $\boldsymbol{q}_N$ closer to $\boldsymbol{z}^*$ and $\boldsymbol{z}^{**}$ (resp.) and increase N.

Splitting of Separatrices and Poincaré–Melnikov Function In Hamiltonian systems, the *Poincaré–Melnikov function* (actually an integral), measures how much the intersecting stable and unstable manifolds of two hyperbolic fixed points split. This kind of function has a long and rich history: Poincaré (1899) introduced it as a way to prove non-integrability in Hamiltonian systems. It has then been used to prove the existence of Chaos (transverse intersections of stable and unstable manifolds often lead to "horseshoe" subsystems), and to estimate the rate of diffusion of orbits in the momentum direction. The discrete, two dimensional case was considered by Easton (1984)???, Gambaudo (1985), Glasser & al. (1989), Delshams & Ramírez-Ros (1996). Here, following Lomeli (1997), we give a formula for a Poincaré–Melnikov function for a higher dimensions symplectic twist map in terms of its generating function. A more general treatment, valid in general cotangent bundles, and which does not assume that the separatrix is a graph over the zero section, is given in Delshams & Ramírez-Ros (1997).

Theorem 37.3 *Let F_0 be an symplectic twist map of $T^*\mathbb{T}^n$ with hyperbolic fixed points $\boldsymbol{z}^* = (\boldsymbol{q}^*, \boldsymbol{p}^*), \boldsymbol{z}^{**} = (\boldsymbol{q}^{**}, \boldsymbol{p}^{**})$ such that $\mathcal{W}^u(\boldsymbol{z}^*) = \mathcal{W}^s(\boldsymbol{z}^{**}) = \mathcal{W}$ is the graph $\boldsymbol{p} = \psi(\boldsymbol{q})$ of a function ψ over some open set. Let S_0 be the generating function of F_0. Consider a perturbation F_ε of F_0 with generating function $S_\varepsilon = S + \varepsilon P$ such that $P(\boldsymbol{q}^*, \boldsymbol{q}^*, 0) = P(\boldsymbol{q}^{**}, \boldsymbol{q}^{**}, 0) = 0$ and $\frac{d}{dq}\big|_{\boldsymbol{q}=\boldsymbol{q}^*} P(\boldsymbol{q}, \boldsymbol{q}, \varepsilon) = 0 = \frac{d}{dq}\big|_{\boldsymbol{q}=\boldsymbol{q}^{**}} P(\boldsymbol{q}, \boldsymbol{q}, \varepsilon)$. Then the function $L : \mathcal{W} \to \mathbb{R}$:*

$$L(\boldsymbol{z}) = \sum_{k \in \mathbb{Z}} P(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}, 0) \quad \text{where } \boldsymbol{q}_k = \pi \circ F^k(\boldsymbol{q}, \psi(\boldsymbol{q})) \tag{37.2}$$

is well defined and differentiable. If L is not constant then, for ε small enough, the (un)stable manifolds of the perturbed fixed points of F_ε split. Their intersection is transverse at nondegenerate critical points of L.

Proof. Work in the covering space $\mathbb{R}2n$ of $T^*\mathbb{T}^n$. Let $\Phi : U \subset \mathbb{R}^n \to \mathbb{R}$ and $\psi = d\Phi$ be such that $Graph(\psi) = \mathcal{W}$. Change coordinates so that $\mathcal{W}$ lies in the zero section: $(\boldsymbol{q}, \boldsymbol{p}') = (\boldsymbol{q}, \boldsymbol{p} - \psi(\boldsymbol{q}))$. If $F_0(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})$, then, in the coordinates $(\boldsymbol{q}, \boldsymbol{p}')$, we have $\boldsymbol{q} = \boldsymbol{q},\ \boldsymbol{p}' = \boldsymbol{p} - \psi(\boldsymbol{q}), \boldsymbol{Q}' = \boldsymbol{Q}, \boldsymbol{P}' = \boldsymbol{P} - \psi(\boldsymbol{Q})$. Thus the generating function becomes:

$$S_{new}(\boldsymbol{q}, \boldsymbol{Q}) = S_{old}(\boldsymbol{q}, \boldsymbol{Q}) + \Phi(\boldsymbol{q}) - \Phi(\boldsymbol{Q}).$$

Note that P remains the same under this change of coordinates, since we only added terms which are independent of ε. For ε small enough, the (un)stable manifolds $\mathcal{W}^u_\varepsilon, \mathcal{W}^s_\varepsilon$ of the perturbed fixed points $\boldsymbol{z}^*_\varepsilon, \boldsymbol{z}^{**}_\varepsilon$ (respectively) will be graphs of the differentials $\psi^{u,s}_\varepsilon = d\phi^{u,s}_\varepsilon$ for some functions $\Phi^{u,s}_\varepsilon$ of the base variable $\boldsymbol{q}$. Clearly, the manifolds $\mathcal{W}^{u,s}_\varepsilon$ split for ε small enough whenever the following *Poincaré–Melnikov* function:

$$M(\boldsymbol{q}) = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (\Phi^u_\varepsilon(\boldsymbol{q}) - \Phi^s_\varepsilon(\boldsymbol{q}))$$

is not constantly zero, and their intersection is transverse if the DM is invertible at the zeros. We will now show that:

$$M(\boldsymbol{q}) = \frac{\partial L}{\partial \boldsymbol{q}}$$

where $L(\boldsymbol{q})$ is the function defined in (37.2) , expressed in our new coordinates. Formula (37.1) gives us an expression of $\Phi^{u,s}_\varepsilon$:

$$\Phi^u_\varepsilon(\boldsymbol{q}) = \sum_{k<0} \left[S_\varepsilon(\boldsymbol{q}^u_k(\varepsilon), \boldsymbol{q}^u_{k+1}(\varepsilon)) - S_\varepsilon(\boldsymbol{q}^{**}, \boldsymbol{q}^{**})\right], \quad \Phi^s_\varepsilon(\boldsymbol{q}) = -\sum_{k\geq 0} \left[S_\varepsilon(\boldsymbol{q}^s_k(\varepsilon), \boldsymbol{q}^s_{k+1}(\varepsilon)) - S_\varepsilon(\boldsymbol{q}^*, \boldsymbol{q}^*)\right]$$

where $\boldsymbol{q}^u_k(\varepsilon)$ (resp $\boldsymbol{q}^s_k(\varepsilon)$) is the $\boldsymbol{q}$ coordinate of $F^k_\varepsilon(\boldsymbol{q}, \psi^u_\varepsilon(\boldsymbol{q}))$ (resp. of $F^k_\varepsilon(\boldsymbol{q}, \psi^s_\varepsilon(\boldsymbol{q}))$). We can change order of differentiation:

$$M(\boldsymbol{q}) = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \frac{\partial}{\partial \boldsymbol{q}}(\psi^u_\varepsilon(\boldsymbol{q}) - \psi^s_\varepsilon(\boldsymbol{q})) = \frac{\partial}{\partial \boldsymbol{q}} \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (\psi^u_\varepsilon(\boldsymbol{q}) - \psi^s_\varepsilon(\boldsymbol{q})),$$

and compute one of these terms:

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{q}} \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \psi^u_\varepsilon(\boldsymbol{q}) \\
&= \frac{\partial}{\partial \boldsymbol{q}} \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \sum_{k<0} \left[S_\varepsilon(\boldsymbol{q}^u_k(\varepsilon), \boldsymbol{q}^u_{k+1}(\varepsilon)) - S_\varepsilon(\boldsymbol{q}^{**}, \boldsymbol{q}^{**})\right] \\
&= \sum_{k<0} \left[\partial_1 S(\boldsymbol{q}^u_k(0), \boldsymbol{q}^u_{k+1}(0)) \frac{\partial}{\partial \varepsilon} \boldsymbol{q}^u_k(\varepsilon) + \partial_2 S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \frac{\partial}{\partial \varepsilon} \boldsymbol{q}^u_{k+1}(\varepsilon) + P(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})\right] \\
&= \sum_{k<0} P(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}),
\end{aligned}$$

where in the last line we took advantage of $\partial_1 S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = 0$: these are the $\boldsymbol{p}$ coordinates of an orbit on the zero section in our new coordinate. In the line before the last, the terms involving $S_\varepsilon(\boldsymbol{q}^{**}, \boldsymbol{q}^{**})$ disappeared because of our assumption on P. The same computation shows that $\frac{\partial}{\partial \boldsymbol{q}} \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} \psi^s_\varepsilon(\boldsymbol{q}) = -\sum_{k\geq 0} P(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$. The proof that $\frac{\partial L}{\partial \boldsymbol{q}} = M(\boldsymbol{q})$ follows. □

Remark 37.4 We have only touched the surface of a vast subject here. Once a Melnikov function is found, one has to be able to show that it is non zero on specific examples. This is usually hard, even in dimension 2. Explicit computations often utilizes the fact that, in good situations, the complexified Melnikov function (think of $\boldsymbol{q}$ as complex in the above) is an elliptic functions. As a result of such computations, one often finds (*eg.* for standard like maps) that the angle of splitting of the separatrices are *exponentially small in the perturbation parameter* ε, making numerical methods inapplicable. We let the reader consult Delshams & Ramírez-Ros (1996b), Delshams & Ramírez-Ros (1998), Glasser & al. (1989), Gelfriech & al. (1994).

Exercise 37.5 a) Prove that the local (un)stable manifold of a hyperbolic fixed point $\boldsymbol{z}^*$ for a symplectic twist map F is a graph over the zero section (*Hint.* use the formula for the differential of F given in 26.5, and the twist condition det $(\partial_{12} S) \neq 0$ to show that the (un)stable subspace of $DF_{\boldsymbol{z}^*}$ cannot have a vertical vector. To do this, expend $\omega_{\boldsymbol{z}^*}(DF\boldsymbol{w}, \boldsymbol{w})$ assuming $\boldsymbol{w} = (0, w)$ and show that necessarily $w = 0$.)
b) Deduce from this that the (un)stable manifolds are graphs of differentials of functions Φ^u, Φ^s defined on a neighborhood of $\pi(\boldsymbol{z}^*)$ in the zero section.

Exercise 37.6 Let F is an exact symplectic map (not necessarily twist) of the cotangent bundle T^*M of some manifold: $F^\lambda - \lambda = dS$ for some function $S : M \to \mathrm{I\!R}$ (λ is the canonical 1 form on T^*M). In Appendix 1 or SG, it is shown that any Hamiltonian map is exact symplectic, and any composition of exact symplectic map is exact symplectic.
a) Show that the (un)stable manifolds $\mathcal{W}^{s,u}$ of a fixed point are *exact Lagrangian* (immersed) submanifolds, *i.e.* $i\mathcal{W}^{s,u}\big|_\lambda = dL$ for some functions $L^{s,u} : \mathcal{W}^{s,u} \to \mathrm{I\!R}$
b) Show that if and $\mathcal{W}$ is an exact Lagrangian manifold invariant under the exact symplectic map F, then:

$$S(z) + constant = L(F(z)) - L(z), \quad \forall p \in \mathcal{W}$$

c) Conclude that $L^u(z^u) = \sum_{k<0} S(F^k(z^u))$ and $L^s(z^s) = -\sum_{k\geq 0} S(F^k(z^s))$.
For more on this approach, see Delshams & Ramírez-Ros (1997).

*. INSTABILITY, TRANSPORT AND DIFFUSION

Part (2) of Birkhoff's Theorem 36.4 is responsible for the name *region of instability* for a region located between two invariant circles, and which does not contain any other invariant circle. This is better understood in the light of the twist maps that appear as normal forms around elliptic fixed points (see Section 28.0): In this example, the lower boundary of the annulus corresponds to the fixed point, and the drifting from the lower boundary to the upper one reflects instability of the fixed point. Mather (1993) and Hall (1989) show that the dynamics in the regions of instability can be quite complicated: given any (infinite) sequence of Aubry-Mather sets in such a region, they find an orbit that shadows it, *i.e.* stays at a prescribed distance from each one for a prescribed amount of time (the transition time is not controlled). In particular, for twist maps of the cylinder without any invariant circles, there exist orbits that are unbounded on the cylinder. To find these orbits, it suffices to take an orbit that shadows an unbounded sequence of Aubry-Mather sets. Note that Slijepčević (1999a) has recently given a proof of these results using the gradient flow of the action methods of GCchapter.

Another approach to instability uses *partial barriers*: invariant sets made of stable and unstable manifolds of hyperbolic periodic orbits or Cantori. The *theory of transport* seeks to study the rate at which points cross these barriers. This theory was initiated by MacKay, Meiss & Percival (1984). The survey Meiss (1992) is beautifully written and encompasses the theory of twist maps of the annulus and transport theory. For other developments, see Rom-Kedar & Wiggins (1990) and Wiggins (1990). MacKay suggested that (the projection in the annulus of) ghost circles could be used as partial barriers.

Mather has announced a striking result for Hamiltonian systems on $T^*\mathbb{T}^2$: For a C^r ($r \geq 2$) generic Riemannian metric g on $\mathbb{T}^2$ and C^r generic potential V periodic in time, the classical Hamiltonian system $H(\boldsymbol{q}, \boldsymbol{p}, t) = \frac{1}{2}\|\boldsymbol{p}\|_g^2 + V(\boldsymbol{q}, t)$ possesses an unbounded orbit. Mather's proof brings together beautifully the constrained variational methods developed in Mather (1993) , the theory of minimal measures of Mather (1991b) as well as hyperbolic techniques. Delshams, de la Llave & Seara (1998)have given recently an alternate proof to this result, using methods of geometric perturbation. Finally, de la Llave just (fall 1999) announced a generalization of this theorem to cotangents bundles of arbitrary compact manifolds. His method uses a generalizations of Fenichel's theory of perturbation of normally hyperbolic sets. Interestingly, the orbits found start at high energy levels, where the system is close to integrable, marking a clear distinction with the two dimensional case where invariant circles would block the escape of orbits. In higher dimensions, KAM tori do not topologically obstruct the passage to higher energy levels.

These results offer a very significant contribution to the problem of diffusion first encountered by Arnold. (???? complete this)

Theorem INVthmkam is 35.1, PBremarkcompactam is 11.4, Section PBsecpbsymp is 12, Theorem PBthmbirkhoff is 36.4, Theorem INVthmgraphmin is 36.2, Exercise INVexoexactstabw is 37.6

7 or HAM

HAMILTONIAN SYSTEMS VS. TWIST MAPS

March 14, 1999

The last section (elliptic f.p.) also appears in SG. Decide where to put it.

In this chapter, we explore the relationship between Hamiltonian systems and symplectic twist maps on cotangent bundles. In the first part of this chapter, we show how to write Hamiltonian systems as compositions of symplectic twist maps. This is instrumental in setting up a simple variational approach to these systems, which is finite dimensional when one searches for periodic orbits. We start in Section 38 with the geodesic flow, which serves as a reference model for Hamiltonian systems: it plays a role similar to that of the integrable map in the twist map theory. In Section 39, we expend our approach to general Hamiltonian or Lagrangian systems satisfying the Legendre condition (which we see as an analog to the twist condition). In Section 3, we show that, whether or not the Legendre condition is satisfied, the time 1 map of a Hamiltonian system may be decomposed into finitely many symplectic twist maps . This method generalises the classical method of broken geodesics of Riemannian geometry. Our main contribution is to make such a method available for Hamiltonian systems that do not satisfy the Legendre condition.

In Section 41, we see how symplectic twist maps also arise from Hamiltonian systems as Poincaré section maps around elliptic periodic orbits. From an opposite perspective, we show in Section 42 that in many cases, a symplectic twist map may be written as the time 1 of a (time dependant) Hamiltonian system. Most of this last section is courtesy of M. Bialy and L. Polterovitch.

38. Case Study: The Geodesic Flow

A. A Few Facts About Riemannian Geometry

Let (M, g) be a compact *Riemannian manifold*. This means that the tangent fibers T_qM are endowed with symmetric, positive definite bilinear forms:

$$(\boldsymbol{v}, \boldsymbol{v}') \mapsto g_{(\boldsymbol{q})}(\boldsymbol{v}, \boldsymbol{v}') \text{ for } \boldsymbol{v}, \boldsymbol{v}' \in T_{\boldsymbol{q}}M$$

varying smoothly with the base point $\boldsymbol{q}$. We will denote the *norm* induced by this metric by $\|\boldsymbol{v}\| := \sqrt{g_{(\boldsymbol{q})}(\boldsymbol{v}, \boldsymbol{v})}$. A curve $\boldsymbol{q}(t)$ in M is a *geodesic* if and only if it is an extremal of the *action* or *energy functional*:

$$A_{t_1}^{t_2}(\gamma) = \int_{t_1}^{t_2} \frac{1}{2} \|\dot{\boldsymbol{q}}\|^2 \, dt.$$

between any two of its points $\boldsymbol{q}(t_1)$ and $\boldsymbol{q}(t_2)$ among all absolutely continuous curves $\beta : [t_1, t_2] \to M$ with same endpoints. Geodesics are usually thought of as length extremals, that is critical points of the functional $\int \frac{1}{2} \|\dot{\boldsymbol{q}}\| \, dt$. But action extremals are length extremals and vice versa (with the difference that action extremals come with a specified parametrization). One usually chooses to compute with the action, since it yields simpler calculations. For more detail on this, as well as a the more abstract definition of geodesic given in terms of a connection see e.g. Milnor (1969) .

The variational problem of finding critical points of A has the Lagrangian

$$L_0(\boldsymbol{q}, \boldsymbol{v}) = \frac{1}{2} g_{(\boldsymbol{q})}(\boldsymbol{v}, \boldsymbol{v}) = \frac{1}{2} \|\dot{\boldsymbol{q}}\|^2 .$$

Following the procedure of Section ??? of Chapter SG, we use the Legendre transform to compute the corresponding Hamiltonian function. In local coordinates $\boldsymbol{q}$ in M, we can write

$$g_{(\boldsymbol{q})}(\boldsymbol{v}, \boldsymbol{v}) = \langle A^{-1}_{(\boldsymbol{q})} \boldsymbol{v}, \boldsymbol{v} \rangle,$$

where $\langle , \rangle$ denotes the dot product in $\mathbb{R}^n$, and $A^{-1}_{(\boldsymbol{q})}$ is a symmetric, positive definite matrix varying smoothly with the base point $\boldsymbol{q}$. With this notation, we have

$$\frac{\partial L_0}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = A^{-1}_{(\boldsymbol{q})} \boldsymbol{v}, \quad \frac{\partial^2 L_0}{\partial \boldsymbol{v}^2} = A^{-1}_{(\boldsymbol{q})}$$

In particular, $\frac{\partial^2 L_0}{\partial \boldsymbol{v}^2}$ is nondegenerate. Hence the Legendre condition is satisfied and the Legendre transformation is, in coordinates:

$$\mathcal{L} : (\boldsymbol{q}, \boldsymbol{v}) \to (\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}, A^{-1}_{(\boldsymbol{q})} \boldsymbol{v})$$

which transforms L_0 into a Hamiltonian H_0:

$$H_0(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p}\boldsymbol{v} - L_0(\boldsymbol{q}, \boldsymbol{v}) = \langle \boldsymbol{p}, A_{(\boldsymbol{q})} \boldsymbol{p} \rangle - \frac{1}{2} \langle A^{-1}_{(\boldsymbol{q})} A_{(\boldsymbol{q})} \boldsymbol{p}, A_{(\boldsymbol{q})} \boldsymbol{p} \rangle = \frac{1}{2} \langle A_{(\boldsymbol{q})} \boldsymbol{p}, \boldsymbol{p} \rangle.$$

This Hamiltonian is a *metric on the cotangent bundle*:

$$H_0(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2} \langle A_{(\boldsymbol{q})} \boldsymbol{p}, \boldsymbol{p} \rangle \stackrel{\text{def}}{=} \frac{1}{2} g^{\#}_{(\boldsymbol{q})}(\boldsymbol{p}, \boldsymbol{p}).$$

We will also denote the norm associated to this metric by $\|\boldsymbol{p}\| = \sqrt{g^{\#}_{(\boldsymbol{q})}(\boldsymbol{p}, \boldsymbol{p})}$. Note that the Legendre transformation is in this case an isometry between the metrics g and $g^{\#}$: in particular, if $(\boldsymbol{q}, \boldsymbol{p}) = \mathcal{L}(\boldsymbol{q}, \boldsymbol{v})$, then $\|\boldsymbol{p}\| = \|\boldsymbol{v}\|$. Hence the Hamiltonian is half of the speed and we retrieve, from conservation of energy in Hamiltonian systems, the fact well known by geometers that extremals of the action are parametrized at constant speed.

The *geodesic flow* is the Hamiltonian flow h_0^t generated by H_0 on T^*M. It is not hard to see that the trajectories of the geodesic flow restricted to an energy level project to the same curves on M as the trajectories in any other energy level: the velocities are are just multiplied by a scalar (See Exercise 38.1). For this reason, one often restricts the geodesic flow to the *unit cotangent bundle* $T_1^*M = \{(\boldsymbol{q}, \boldsymbol{p}) \in T^*M \mid \|\boldsymbol{p}\| = 1\}$. Traditionally, geometers use the term geodesic flow to denote the conjugate $\mathcal{L}^{-1} h_0^t \mathcal{L}$ on TM of this Hamiltonian flow, as restricted to the unit tangent bundle. Remember that projections of trajectories of a Hamiltonian flow associated to a Lagrangian satisfying the Legendre condition are extremals of the action of the Lagrangian, and vice versa. (See Chapter SG, Section ???). In the present case, if $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ a trajectory of the geodesic flow, then $\boldsymbol{q}(t)$ is a geodesic. Conversely, if $\boldsymbol{q}(t)$ is a geodesic, it is the projection on M of the solution $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ of the geodesic flow with initial condition $(\boldsymbol{q}_0, \boldsymbol{p}_0) = (\boldsymbol{q}(0), A^{-1}\dot{\boldsymbol{q}}(0))$.

We now want to establish a fundamental result of Riemmannian geometry, which we will rephrase in the next subsection by saying that the time t of the geodesic flow is a symplectic twist map . The *exponential map* is defined by:

$$exp_{\boldsymbol{q}_0}(t\boldsymbol{v}) = \boldsymbol{q}(t),$$

where $\boldsymbol{q}(t)$ is the geodesic such that $\dot{\boldsymbol{q}}(0) = \boldsymbol{v}$. Note that any geodesic can be written in this exponential notation. In terms of the geodesic flow, $exp_{\boldsymbol{q}_0}(t\boldsymbol{v}) = \pi \circ h_0^t \circ \mathcal{L}(\boldsymbol{q}_0, \boldsymbol{v})$, where $\pi : T^*M \mapsto M$ is the canonical projection.

Theorem 38.1 *The map* $Exp : TM \to M \times M$

$$(\boldsymbol{q}, \boldsymbol{v}) \mapsto (\boldsymbol{q}, \boldsymbol{Q}) \stackrel{\text{def}}{=} (\boldsymbol{q}, exp_{\boldsymbol{q}}(\boldsymbol{v})) \tag{38.1}$$

defines a diffeomorphism between a neighborhood of the 0*–section in* TM *and some neighborhood of the diagonal in* $M \times M$*. Moreover, for* $(\boldsymbol{q}, \boldsymbol{v})$ *in that neighborhood:*

$$\text{Dis}(\boldsymbol{q}, exp_{\boldsymbol{q}}(\boldsymbol{v})) = \|\boldsymbol{v}\|$$

One way to paraphrase this theorem is by saying that, any two closeby points are joined by a unique, short enough, geodesic segment.

Proof. By definition, $exp_{\boldsymbol{q}}(0) = \boldsymbol{q}$ and $\frac{d}{ds} exp_{\boldsymbol{q}}(s\boldsymbol{v}) = \boldsymbol{v}$ at $s = 0$,

$$DExp\big|_{(\boldsymbol{q},0)} = \begin{pmatrix} Id & Id \\ 0 & Id \end{pmatrix},$$

whose determinant is 1. Hence, Exp is a local diffeomorphism around each point of a compact neighborhood of the 0-section. In particular we can assume that there is an ϵ such that Exp is a diffeomorphism of an ϵ ball in TM around $(\boldsymbol{q}, 0)$ and a ball in $M \times M$ around $(\boldsymbol{q}, \boldsymbol{q})$, where ϵ is independant of $\boldsymbol{q}$.

We now show that Exp is an embedding when restricted to $U_\epsilon = \{(\boldsymbol{q}, \boldsymbol{v}) \in TM \mid \|\boldsymbol{v}\| \le \epsilon\}$, where ϵ is as above. It is enough to check the injectivity. Let two elements in U_ϵ have the same image under Exp. Since the first factor of Exp gives the base point, this can only occur if they are in the same fiber of U_ϵ. But, by our choice of U_ϵ this implies these elements are the same.

Finally, we show that $\text{Dis}(\boldsymbol{q}, exp_{\boldsymbol{q}}(\boldsymbol{v})) = \|\boldsymbol{v}\|$ whenever $\|\boldsymbol{v}\| \le \epsilon$. We remind the reader that the *distance* $\text{Dis}(\boldsymbol{q}, \boldsymbol{Q})$ between two points $\boldsymbol{q}$, and $\boldsymbol{Q}$ in a compact Riemannian manifold is given by the length of the shortest path between $\boldsymbol{q}$ and $\boldsymbol{Q}$. As a length minimizer, the shortest path is also an action minimizer, and hence a geodesic. Since Exp is an embedding of U_ϵ in $M \times M$, exp is 1 to 1 on $U_\epsilon \cap T_{\boldsymbol{q}}M$ and the unique geodesic that joins $\boldsymbol{q}$ and $exp_{\boldsymbol{q}}(\boldsymbol{v})$ in $exp(U_\epsilon \cap T_{\boldsymbol{q}}M)$ is the curve $t \mapsto \boldsymbol{q}(t) = exp_{\boldsymbol{q}}(t\boldsymbol{v})$. The length of this curve is $\int_0^1 \|\dot{\boldsymbol{q}}\|\, dt = \int_0^1 \|\boldsymbol{v}\|\, dt = \|\boldsymbol{v}\|$ (see Exercise 38.1 c)). The only way our formula may fail is if there were a shorter geodesic joining $\boldsymbol{q}$ and $exp_{\boldsymbol{q}}(\boldsymbol{v})$ not in $exp(U_\epsilon \cap T_{\boldsymbol{q}}M)$). But this is impossible since this geodesic would be of the form $\exp_{\boldsymbol{q}}(t\boldsymbol{w}), t \in [0, 1]$ with length $\|\boldsymbol{w}\| > \epsilon$.

□

Exercise 38.1 a) Check that, in local coordinates, Hamilton's equations for the geodesic flow write:

$$\begin{aligned}\dot{q} &= A_{(q)}\boldsymbol{p} \\ \dot{p} &= -\left\langle \frac{\partial A_{(q)}}{\partial \boldsymbol{q}}\boldsymbol{p}, \boldsymbol{p}\right\rangle\end{aligned} \tag{38.2}$$

b) Verify that $h_0^{st}(\boldsymbol{q},\boldsymbol{p}) = h_0^t(\boldsymbol{q}, s\boldsymbol{p})$. (*Hint.* if $(\boldsymbol{q}(t),\boldsymbol{p}(t))$ is a trajectory of the geodesic flow, then $(\boldsymbol{q}(st), s\boldsymbol{p}(st))$ is also a trajectory).
c) Show that if $\boldsymbol{q}(t) = exp_{\boldsymbol{q}_0}(t\boldsymbol{v})$, $\|\dot{\boldsymbol{q}}(t)\| = \|\boldsymbol{v}\|$ for all t.

Exercise 38.2 Show that the completely integrable twist map $(x, y) \mapsto (x + y, y)$ is the time 1 map of the geodesic flow on the "flat" circle, *i.e.* the circle given the euclidean metric $g_{(x)}(v, v) = v^2$.

B. The Geodesic Flow As A Twist Map

Theorem 38.1 is the key to the following:

Proposition 38.2 *The time 1 map h_0^1 of the geodesic flow with Hamiltonian $H_0(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{p}\|^2$ is a symplectic twist map on $U_\epsilon = \{(\boldsymbol{q},\boldsymbol{p}) \in T^*M \mid \mid \|\boldsymbol{p}\| \le \epsilon\}$, for ϵ small enough. More generally, given any $R > 0$, there is an $t_0 > 0$ (or given any t_0 there is an R) such that $h_0^t, t \in [-t_0, t_0]$, is a symplectic twist map on the set $U_R = \{(\boldsymbol{q},\boldsymbol{p}) \mid \mid \|\boldsymbol{p}\| \le R\}$. The generating function of h_0^t is given by $S(\boldsymbol{q},\boldsymbol{Q}) = \frac{t}{2}\mathrm{Dis}^2(\boldsymbol{q},\boldsymbol{Q})$.*

Proof. Since h_0^1 is a Hamiltonian map, it is exact symplectic. Define $Exp^\# = Exp \circ \mathcal{L}^{-1}$. By Theorem 38.1, $Exp^\#$ is a diffeomorphism between $U_\epsilon = \{(\boldsymbol{q},\boldsymbol{p}) \mid \|\boldsymbol{p}\| = \epsilon\}$ and a neighborhood of the diagonal in $M \times M$. But $Exp^\#(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}, \boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p}))$, where $\boldsymbol{Q} = \pi \circ h_0^1(\boldsymbol{q},\boldsymbol{p})$. Hence h_0^1 is a symplectic twist map on U_ϵ, and $\psi_{h_0^1} = Exp^\#$. The more general statement derives from the fact that $Exp^\#(\boldsymbol{q}, t\boldsymbol{p}) = (\boldsymbol{q}, \boldsymbol{q}(t))$, where $h_0^t(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}(t), \boldsymbol{p}(t))$.

We now show that $\frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q},\boldsymbol{Q})$ is the generating function of h_0^1 when it is a symplectic twist map on a domain U (the proof for h_0^t is identical). Since h_0^1 is a Hamiltonian map,

$$(h_0^1)^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = dS, \quad \text{with} \quad S(\boldsymbol{q},\boldsymbol{p}) = \int_\gamma \boldsymbol{p}d\boldsymbol{q} - H_0 dt$$

where γ is the curve $h_0^t(\boldsymbol{q},\boldsymbol{p}),\ t \in [0,1]$ (see Theorem ??? in Chapter SG). We now need to show that S, expressed as a function of $\boldsymbol{q},\boldsymbol{Q}$ is the one advertised. In this particular case, since $\dot{\boldsymbol{q}} = A_{(q)}\boldsymbol{p}$ (see Exercise 38.1) and $H_0 = \frac{1}{2}\langle A_{(q)}\boldsymbol{p},\boldsymbol{p}\rangle = \frac{1}{2}\|\boldsymbol{p}\|^2$, the integral simplifies:

$$\int_\gamma \boldsymbol{p}d\boldsymbol{q} - H_0 dt = \int_0^1 \frac{1}{2}\langle A_{(q)}\boldsymbol{p},\boldsymbol{p}\rangle - \frac{1}{2}\|\boldsymbol{p}(t)\|^2\, dt = \int_0^1 \frac{1}{2}\|\boldsymbol{p}(t)\|^2\, dt$$

But the integrand is H_0, which is constant along γ. Hence, using Theorem 38.1 , and the fact that $\mathcal{L}$ is an isometry, we get:

$$S(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{p}\|^2 = \frac{1}{2}\|\dot{\boldsymbol{v}}\|^2 = \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{Q}(\boldsymbol{q},\boldsymbol{p})),$$

where $(\boldsymbol{q},\boldsymbol{v}) = \mathcal{L}^{-1}(\boldsymbol{q},\boldsymbol{p})$. This makes S the advertised differentiable function of $\boldsymbol{q}$ and $\boldsymbol{Q}$ whenever $(\boldsymbol{q},\boldsymbol{p}) \mapsto (\boldsymbol{q},\boldsymbol{Q})$ is a diffeomorphism. □

Remark 38.3 As a simple example of what makes h_0^1 cease to be a twist map when the domain U is extended too far, take M to be the unit circle with the arclength metric. In a chart $\theta \in (-\epsilon, 2\pi - \epsilon)$, we have:

$$\mathrm{Dis}(0,\theta) = \begin{cases} \theta & \text{when} \quad \theta \le \pi \\ 2\pi - \theta & \text{when} \quad \theta > \pi \end{cases}$$

As a result, the left derivative of $\frac{1}{2}\mathrm{Dis}^2(0,\theta)$ is π, whereas the right derivative is $-\pi$: the function Dis is not differentiable at this point.

The following will be instrumental in the proof of Theorem 31.1. ???Put it there instead???

Corollary 38.4 *Let $h_0^s(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{Q}_s, \boldsymbol{P}_s)$ be the time s of the geodesic flow, then:*

$$\partial_1 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_s) = -sign(s).\frac{\boldsymbol{p}}{\|\boldsymbol{p}\|} \quad and \quad \partial_2 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_s) = sign(s).\frac{\boldsymbol{P}_s}{\|\boldsymbol{P}_s\|} \tag{38.3}$$

Proof. From Proposition 38.2, we get:

$$-\boldsymbol{p} = \partial_1 \frac{1}{2}\mathrm{Dis}^2(\boldsymbol{q}, \boldsymbol{Q}_1) = \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_1)\partial_1 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_1) = \|\boldsymbol{p}\|\, \partial_1 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_1)$$

which proves $\partial_1 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_1) = -\frac{\boldsymbol{p}}{\|\boldsymbol{p}\|}$. Using $\boldsymbol{Q}_s = \pi \circ h_0^1(\boldsymbol{q}, s\boldsymbol{p})$, one may replace $\boldsymbol{p}$ by $s\boldsymbol{p}$ in the previous computation to prove the first equality. For the second equality, the fact that $\mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_s) = \mathrm{Dis}(\boldsymbol{Q}_s, \boldsymbol{q})$, that $\boldsymbol{q} = \pi \circ h_0^1(\boldsymbol{Q}_s, -s\boldsymbol{P}_s)$ (see Exercise 38.2) and the first equality, enables us to write:

$$\partial_2 \mathrm{Dis}(\boldsymbol{q}, \boldsymbol{Q}_s) = \partial_1 \mathrm{Dis}(\boldsymbol{Q}_s, \boldsymbol{q}) = sign(s).\frac{\boldsymbol{P}_s}{\|\boldsymbol{P}_s\|}$$

□

C. The Method of Broken Geodesics

We now draw the correspondence between the variational methods provided by symplectic twist maps and the classical method of broken geodesics, originally due to Birkhoff (???: check Milnor). As before, let h_0^1 be the time 1 map[7] of the geodesic flow with Hamiltonian H_0. Fix some neighborhood U of the zero section in T^*M. Proposition 38.2implies that if we decompose $h_0^1 = (h_0^{\frac{1}{N}})^N$, then for N big enough each $h_0^{\frac{1}{N}}$ is a symplectic twist map in U. As a result, periodic orbits of period 1 for the geodesic flow H_0, *i.e.* fixed points of h_0^1 are given by the critical points of:

$$W(\overline{\boldsymbol{q}}) = \sum_{k=1}^{N} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}), \quad \text{with} \quad \boldsymbol{q}_{N+1} = \boldsymbol{q}_1,$$

where $\overline{\boldsymbol{q}}$ belong to set $X_N(U)$ of sequences in M such that $(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \in \psi(U)$, where $\psi = \psi_{h_0^{\frac{1}{N}}}$ We now show that W is the action of a *broken geodesic*. Since $h_0^{\frac{1}{N}}$ is a symplectic twist map, the twist condition implies that, given $(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ in $\psi(U)$, there is a unique $(\boldsymbol{p}_k, \boldsymbol{P}_k)$ such that $h_0^{\frac{1}{N}}(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{P}_k)$, i.e., there is exactly one trajectory $c_k\colon [\frac{k}{N}, \frac{k+1}{N}] \to T^*\tilde{M}$ of the geodesic flow that joins $(\boldsymbol{q}_k, \boldsymbol{p}_k)$ to $(\boldsymbol{q}_{k+1}, \boldsymbol{P}_k)$. The projection $\pi(c_k)$ on M is a geodesic, parametrized at constant speed equal to the norm of $\boldsymbol{p}_k$. As we

[7] The following discussion remains valid if we replace 1 by any time T.

have seen in the proof of Proposition 38.3, $S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ is the action of c_k: $S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = \int_{c_k} \boldsymbol{p}d\boldsymbol{q} - Hdt$. Hence W, the sum of these actions, is the action of the curve C obtained by the concatenation of the c_k's. C is "broken", $i.e.$ has a corner at the point $\boldsymbol{q}_k$ whenever $\boldsymbol{P}_{k-1} \neq \boldsymbol{p}_k$: via the Legendre transformation, $\boldsymbol{P}_{k-1}$ and $\boldsymbol{p}_k$ correspond to the left derivative and right derivative of the curve C at $\boldsymbol{q}_k$.

If $\overline{\boldsymbol{q}}$ is a critical point of W, $\boldsymbol{P}_k = \boldsymbol{p}_{k+1}$, and thus the left and right derivatives coincide: in this case C is a closed, smooth geodesic.

In conclusion, the function $W(\overline{\boldsymbol{q}})$ can be interpreted as the restriction of the action functional $A(c)$ to a *finite dimensional* subspace (the space of curves C arising from elements of $X_N(U)$, which is homeomorphic to $X_N(U)$) in the (infinite dimensional) loop space of T^*M. One can further justify this method by showing that the finite dimensional space $X_N(U)$ is a deformation retract[(8)] of a subset of the loop space and that it contains all the critical loops of that subset. This was Morse's way to study the topology of the loop space (see Milnor (1969) , 16). Conversely, and this is the point of view in this book (and more generally that of symplectic topology), knowing the topology of certain subsets of the loop space, one can gain information about the dynamics of the geodesic flow or, as we will see, of many Hamiltonian systems. (Part of this in the Intro???)

39. Decomposition Of Hamiltonian Maps Into Twist Maps

A. Legendre Condition vs. Twist Condition

In this subsection, we generalize Theorem 38.2 by proving that Hamiltonian maps satisfying the Legendre condition are symplectic twist maps, provided appropriate restrictions on the domain of the map. We then reformulate this result in the Lagrangian setting, giving a generalization of the fundamental Theorem 38.1. In the next subsection, we focus on $T^*\mathbb{T}^n$, where, given further conditions on the Hamiltonian, we extend the domain of these symplectic twist maps to the whole space.

Remember that Hamiltonian maps, which are time t maps of Hamiltonian systems, are exact symplectic (Theorem SGhamexactsymp) and, through the flow, isotopic to Id. Therefore, to show that a certain Hamiltonian map is a symplectic twist map, we need only check the twist condition. Clearly, not all Hamiltonian maps satisfy it. Take $F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{p})$ on the cotangent bundle of the torus, for example: it is the time one of $H(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{m}.\boldsymbol{p}$, and it is definitely not twist. Here is a heuristic argument, which appeared in Moser (1986) in the context of twist maps, to guide us in our search of the twist condition for Hamiltonian maps. The Taylor series with respect to ϵ of the time ϵ map of a Hamiltonian system with Hamiltonian H is:

$$\begin{aligned} \boldsymbol{q}(\epsilon) &= \boldsymbol{q}(0) + \epsilon.H_{\boldsymbol{p}} + o(\epsilon^2) \\ \boldsymbol{p}(\epsilon) &= \boldsymbol{p}(0) - \epsilon.H_{\boldsymbol{q}} + o(\epsilon^2) \end{aligned}.$$

Thus, up to order ϵ^2, $\partial \boldsymbol{q}(\epsilon)/\partial \boldsymbol{p}(0) = \epsilon.H_{\boldsymbol{pp}}$. This shows that *whenever $H_{\boldsymbol{pp}}$ is nondegenerate, the time ϵ map is a symplectic twist map* in some neighborhood of $\boldsymbol{q}(0), \boldsymbol{p}(0)$. The problem is to extend this argument to given regions of the cotangent bundle: the term $o(\epsilon^2)$ might get large as the initial condition varies.

We now present a rigorous version of this argument, valid on compact sets of cotangent bundles of arbitrary compact manifolds. We say that a Hamiltonian $H : T^*M \times \mathbb{R} \to \mathbb{R}$ satisfies the *global Legendre condition* if the map:

[8] This retraction can be obtained by a piecewise curve shortening method.

$$(39.1) \qquad \boldsymbol{p} \mapsto H_p(\boldsymbol{q}, \boldsymbol{p}, t)$$

is a diffeomorphism from $T^*_{\boldsymbol{q}}M \mapsto T_{\boldsymbol{q}}M$ for each $\boldsymbol{q}$ and t. We will say that H satisfies the *Legendre embedding condition* if the map $\boldsymbol{p} \mapsto H_{\boldsymbol{p}}$ is an embedding (*i.e.* a 1–1, local diffeomorphism). Note that, although we have written it in a chart of conjugate coordinates in T^*M, this condition is coordinate independant (prove this!).

Examples 39.1 We give two classes of examples. In the first one, the Hamiltonian is *not* assumed to be convex.

Let $H(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\langle A_{(\boldsymbol{q},t)}\boldsymbol{p}, \boldsymbol{p}\rangle + V(\boldsymbol{q}, t)$ and $\det\ A_{(\boldsymbol{q},t)} \neq 0$, then H satisfies (39.1). This is simply because $\boldsymbol{p} \mapsto H_{\boldsymbol{p}} = A_{(\boldsymbol{q},t)}\boldsymbol{p}$ is linear and nonsingular. Note that no convexity is assumed here, only nondegeneracy of $H_{\boldsymbol{pp}}$ (and its independance of $\boldsymbol{p}$).

Less trivially, if $H_{\boldsymbol{pp}}(\boldsymbol{q}, \boldsymbol{p}, t)$ is definite positive, and its smallest eigenvalue is uniformly bounded below by a strictly positive constant, then H statisfies the global Legendre condition. This is a direct consequence of Lemma STMdiffeo.

If we remove the lower bound on the smallest eigenvalue, one can show (see Exercise 39.1) that the map $\boldsymbol{p} \mapsto H_{\boldsymbol{p}}$ is not necessarily a diffeomorphism any more, but remains an embedding and thus H satisfies the Legendre embedding condition.

Such an embedding condition, and a version of Theorem 39.2 , are also satisfied if $H_{\boldsymbol{pp}}$ is positive on a compact set U invariant under the flow (see Exercise 39.2).

Theorem 39.2 *Let M be a compact, smooth manifold and $H : T^*M \times \mathbb{R}$ be a smooth Hamiltonian function which satisfies either the global Legendre condition (39.1) or the Legendre embedding condition. Then, given any compact neighborhood U in T^*M and starting time a, there exists $\epsilon_0 > 0$ (depending on U) such that, for all $\epsilon < \epsilon_0$ the time ϵ map of the Hamiltonian flow of H is a symplectic twist map on U.*

Proof of Theorem 39.1 Choose a Riemannian metric g on M. Define the compact ball bundles:

$$U(K) = \{(\boldsymbol{q}, \boldsymbol{p}) \in T^*M \mid \|\boldsymbol{p}\| \leq K\}.$$

The nested union of these sets covers T^*M. Hence any compact set U is contained in a $U(K)$ for some K large enough, and we may restrict the proof of the theorem to the case $U = U(K)$. Since the Hamiltonian vector field of H is uniformly Lipschitz on compact sets, there is a time T such that the Hamiltonian flow $h_a^{a+t}(\boldsymbol{z})$ of H is defined on the interval $t \in [0, T]$ whenever $\boldsymbol{z} \in U(K)$.

In the rest of the section, we fix a and abreviate h_a^{a+t} by h^t (the time t of the flow with starting time a.

By continuity of the flow, $h^{[0,T]}(U(K))$ is a compact set. We now show that we can work in appropriately chosen charts of T^*M. Since M is compact, we can find a real $r > 0$ such that T^*M is trivial above each ball of radius $2r$ in M. (Indeed, there exist such a ball around each point. If one had a sequence of points whose corresponding maximum such r converged to zero, a limit point of this sequence would not have a trivializing neighborhood, a contradiction). Take a finite covering $\{B_i\}$ of M by balls of radius r, and let B'_i

be the ball of radius $2r$ with same center as B_i. Choose $\epsilon_3 < T$ such that $\pi \circ h^{[0,\epsilon_3]}(\pi^{-1}(B_i)) \subset B_i'$. Such an ϵ_3 exists since there are finitely many B_i's and since the flow is continuous. From now on, we may work in any of the charts $\pi^{-1}(B_i) \simeq B_i \times \mathbb{R}^n$, and know that for the time interval $[0, \epsilon_3]$, we will remain in the charts $\pi^{-1}(B_i') \simeq B_i' \times \mathbb{R}^n$. We let $(\boldsymbol{q}, \boldsymbol{p})$ denote the conjugate coordinates in these charts.

Let $\epsilon < \epsilon_3$ and write $h^\epsilon(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}(\epsilon), \boldsymbol{p}(\epsilon))$. Consider the map $\psi_{h^\epsilon} : (\boldsymbol{q}, \boldsymbol{p}) \mapsto (\boldsymbol{q}, \boldsymbol{q}(\epsilon))$. We need to show that ψ_{h^ϵ} is an embedding of $U(K)$ in $M \times M$. By compactness, it suffices to show that ψ_{h^ϵ} is a local diffeomorphism which is 1–1 on $U(K)$. Write the second order Taylor formula for $\boldsymbol{q}(\epsilon)$ with respect to ϵ (this is a smooth function since the flow is smooth):

$$\boldsymbol{q}(\epsilon) = \boldsymbol{q} + \epsilon H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}, a) + \epsilon^2 R(\boldsymbol{q}, \boldsymbol{p}, \epsilon).$$

The smoothness of the Hamiltonian flow garantees that R is smooth in all its variables. Indeed, its precise expression is (see Lang (1983) , p. 116):

$$R(\boldsymbol{q}, \boldsymbol{p}, \epsilon) = \int_0^1 (1-t) \frac{\partial h^{t\epsilon}(\boldsymbol{q}, \boldsymbol{p})}{\partial t} dt$$

and the integrand is smooth since the flow is. The differential of ψ_{h^ϵ} with respect to $(\boldsymbol{q}, \boldsymbol{p})$ is of the form:

$$D\psi_{h^\epsilon}(\boldsymbol{q}, \boldsymbol{p}) = \begin{pmatrix} Id & 0 \\ * & A \end{pmatrix}, \quad A = \epsilon H_{\boldsymbol{pp}}(\boldsymbol{q}, \boldsymbol{p}, a) + \epsilon^2 R_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}, \epsilon).$$

Since $\det\, H_{\boldsymbol{pp}} \neq 0$ by the Legendre condition and since $R_{\boldsymbol{p}}$ is continuous and hence bounded on the compact set $U(K) \times [0, \epsilon_3]$, there exists ϵ_2 in $(0, \epsilon_3]$ such that $\det\, D\psi_{h^\epsilon} = \det\, A \neq 0$ on $U(K) \times (0, \epsilon_2]$ (we have used the fact that there are finitely many of our charts B_i covering $U(K)$). Hence ψ_{h^ϵ} is a local diffeomorphism for all $\epsilon \in (0, \epsilon_2]$. We now show that, by maybe shrinking further the interval of ϵ, ψ_{h^ϵ} is one to one on $U(K)$. Suppose not and $\psi_{h^\epsilon}(\boldsymbol{q}, \boldsymbol{p}) = \psi_{h^\epsilon}(\boldsymbol{q}', \boldsymbol{p}')$ for some $(\boldsymbol{q}, \boldsymbol{p}), (\boldsymbol{q}', \boldsymbol{p}') \in U(K)$. The definition of ψ_{h^ϵ} immediately implies that $\boldsymbol{q} = \boldsymbol{q}'$. Also, since ψ_{h^ϵ} is a local diffeomorphism on $U(K)$, we can assume that $\|\boldsymbol{p} - \boldsymbol{p}'\| > \delta$ for some $\delta > 0$. Using Taylor's formula, we have:

$$\boldsymbol{q}(\epsilon) - \boldsymbol{q}'(\epsilon) = \epsilon(H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}, a) - H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}', a)) + \epsilon^2(R(\boldsymbol{q}, \boldsymbol{p}, \epsilon) - R(\boldsymbol{q}, \boldsymbol{p}', \epsilon)).$$

Define the compact set $P(K) := \{(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{q}, \boldsymbol{p}\prime) \in U(K) \times U(K) \mid \|\boldsymbol{p} - \boldsymbol{p}\prime\| \geq \delta\}$. Since $\boldsymbol{p} \mapsto H_{\boldsymbol{p}}$ is a diffeomorphism, the continuous function $\|H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}, a) - H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}', a)\|$ is bounded below by some $K_1 > 0$ on $P(K)$. The continuous function $(\boldsymbol{q}, \boldsymbol{p}, \epsilon) \mapsto \|R(\boldsymbol{q}, \boldsymbol{p}, \epsilon) - R(\boldsymbol{q}, \boldsymbol{p}', \epsilon)\|$ is bounded, say by K_2, on $P(K) \times [0, \epsilon_2]$ and hence

$$\|\boldsymbol{q}(\epsilon) - \boldsymbol{q}'(\epsilon)\| \geq (\epsilon K_1 - \epsilon^2 K_2) > 0$$

whenever $\epsilon \in (0, \epsilon_1]$ and ϵ_1 is small enough. Now choosing $\epsilon_0 = \min\{\epsilon_1, \epsilon_2\}$ finishes the proof of the theorem. □

The following proposition, which is a reformulation of Theorem 39.2 in Lagrangian terms, is a generalization of the fundamental Theorem 38.1 . It garantees the existence and *uniqueness* of Euler-Lagrange solutions between any two closeby points. A time that the solution is traversed has to be specified within a compact interval. In Chapter MIN, we will encounter Tonelli's theorem which implies, for fiber convex Lagrangian systems, that these solutions can also be assumed to be action minimizers.

Proposition 39.3 *Let M be a compact manifold and $L : TM \times \mathbb{R} \to \mathbb{R}$ be a Lagrangian function satisfying the global Legendre condition: $\boldsymbol{v} \mapsto L_{\boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}, t)$ is a diffeomorphism. Then, for all starting time a and bound on the velocity K there exists an interval of time $[a, a + \epsilon_0]$ such that, for all $\epsilon < \epsilon_0$, there exists a neighborhood $\mathcal{O}$ of the diagonal in $M \times M$ such that whenever $(\boldsymbol{q}, \boldsymbol{Q}) \subset \mathcal{O}$, there exists a unique solution $\boldsymbol{q}(t)$ of the Euler-Lagrange equations such that $\boldsymbol{q} = \boldsymbol{q}(a)$, $\boldsymbol{Q} = \boldsymbol{q}(a+\epsilon)$ and $\|\dot{\boldsymbol{q}}(a)\| \leq K$.*

Remark 39.4 Note that, in the case of the geodesic flow, the curves joining the same points $\boldsymbol{q}, \boldsymbol{Q}$ in different time intervals in this proposition are geometrically all the same geodesic, traversed at different speeds. The dependence on the time interval chosen and the speed chosen of the geometric solutions of the Euler-Lagrange equations is one of the main difference, and source of confusion, when trying to generalise notions of Riemannian geometry to Lagrangian mechanics.

Proof. The Legendre condition enables us to define the Legendre transform $\mathcal{L} : (\boldsymbol{q}, \boldsymbol{v}) \to (\boldsymbol{q}, \boldsymbol{p} = L_{\boldsymbol{v}})$ and the Hamiltonian function $H(\boldsymbol{q}, \boldsymbol{p}, t) = \boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$, where it is understood that $\dot{\boldsymbol{q}} = \dot{\boldsymbol{q}} \circ \mathcal{L}^{-1}(\boldsymbol{q}, \boldsymbol{p})$ (see Section hamsys ??? in SG). H satisfies the global Legendre condition and $\mathcal{L}^{-1}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}, H_{\boldsymbol{p}})$ (see Remark ???), In particular Theorem 39.3applies to the Hamiltonian H. Let

$$U = V(K) = \{(\boldsymbol{q}, \boldsymbol{p}) \mid \|H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}, a)\| \leq K\}.$$

This set is compact since it corresponds, under the Legendre transformation, to

$$\mathcal{L}^{-1}(V(K)) = \{(\boldsymbol{q}, \dot{\boldsymbol{q}}) \mid \|\dot{\boldsymbol{q}}(a)\| \leq K\}$$

in the tangent bundle. Theorem 39.3 tells us that, for all $\epsilon \in (0, \epsilon_0]$ with ϵ_0 small enough, the map h^ϵ is a symplectic twist map on $V(K)$. Define

$$\mathcal{O} = \psi_{h^\epsilon}(V(K)).$$

We now show, maybe by decreasing ϵ_0, that $\mathcal{O}$ is a neighborhood of the diagonal in $M \times M$. Let $V_{\boldsymbol{q}}(K) = \pi^{-1}(\boldsymbol{q}) \cap V(K)$ and write $h^t(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}(t), \boldsymbol{p}(t))$ where, as before, h^t denotes h_a^{a+t}. The curve $\boldsymbol{q}(t)$ is a solution of the Euler-Lagrange equation satisfying $\boldsymbol{q} = \boldsymbol{q}(a)$ and if $(\boldsymbol{q}, \boldsymbol{p}) \in V_{\boldsymbol{q}}(K)$, then $\|\dot{\boldsymbol{q}}(a)\| = \|H_{\boldsymbol{p}}\| \leq K$. As in the proof of Theorem 39.2, we write the Taylor approximation of the solution:

$$\pi \circ h^\epsilon(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{q}(\epsilon) = \boldsymbol{q} + \epsilon H_{\boldsymbol{p}} + \epsilon^2 R(\boldsymbol{q}, \boldsymbol{p}, \epsilon).$$

At first order in ϵ, the image of $V_{\boldsymbol{q}}(K)$ under $\pi \circ h^\epsilon$ is $\{\boldsymbol{q} + \epsilon H_{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}) \mid (\boldsymbol{q}, \boldsymbol{p}) \in V_{\boldsymbol{q}}(K)\}$, which is a solid ball centered at $\boldsymbol{q}$. When adding the second order term $\epsilon^2 R$, $\boldsymbol{q}$ will still be in $\pi \circ h^\epsilon(V_{\boldsymbol{q}}(K))$, provided that ϵ is small enough. By compactness ϵ can be chosen to work for all $\boldsymbol{q}$. Thus $(\boldsymbol{q}, \boldsymbol{q}) \in h^\epsilon(V(K)) = \mathcal{O}$ for all $\boldsymbol{q} \in M$, as claimed.

The rest of the proof is a pure translation of the statements of Theorem 39.2: by construction, if $(\boldsymbol{q}, \boldsymbol{Q}) \in \mathcal{O}$, then $(\boldsymbol{q}, \boldsymbol{Q}) = (\boldsymbol{q}, \boldsymbol{q}(\epsilon))$ where $\boldsymbol{q}(t) = \pi \circ h^t(\boldsymbol{q}, \boldsymbol{p})$ and $(\boldsymbol{q}, \boldsymbol{p}) \in V(K)$. Hence $\boldsymbol{q}(t)$ is a solution to the Euler-Lagrange equation starting at $\boldsymbol{q}$ at time a, landing on $\boldsymbol{Q}$ at time $a+\epsilon$. Moreover, since $(\boldsymbol{q}, \boldsymbol{p}) \in V(K)$, $\|\dot{\boldsymbol{q}}(a)\| = \|H_p(\boldsymbol{q}, \boldsymbol{p}, a)\| \leq K$. Finally, this solution is unique. Otherwise, by the uniqueness of solutions of O.D.E.'s, there would be $\boldsymbol{p} \neq \boldsymbol{p}'$ such that $\pi \circ h^\epsilon(\boldsymbol{q}, \boldsymbol{p}) = \pi \circ h^\epsilon(\boldsymbol{q}, \boldsymbol{p}')$, a contradiction to the twist condition. □

Exercise 39.1 Show that a C^1 map $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ which satisfies $\langle Df_x \cdot v, v\rangle > 0$ for all v and x in $\mathbb{R}^n$ is an embedding, *i.e.* it is injective with continuous and differentiable inverse. Deduce that a Hamiltonian such that H_{pp} is positive definite satisfies the Legendre embedding condition.

Exercise 39.2 Let U be a compact region which is invariant under the flow of a Hamiltonian H. Assume also that H_{pp} is positive definite on U. Show that the time t map is a symplectic twist map for all $t > 0$ sufficiently small. (*Hint.* First prove, as in the previous exercise, that $\boldsymbol{p} \mapsto H_{\boldsymbol{p}}$ is an embedding of $T^*_{\boldsymbol{q}}M \cap U$ for each $\boldsymbol{q}$. Then adapt the proof of Theorem 39.2).

B. The Case of the Torus

When the configuration manifold is $\mathbb{T}^n$, there is hope to show that the time t maps of a Hamiltonian system is a symplectic twist map on the whole cotangent bundle. We present here some condition under which this is true. No doubt one could find other, even weaker conditions as well.

Assumption 1 (Uniform opticity)

$H(\boldsymbol{q}, \boldsymbol{p}, t) = H_t(\boldsymbol{z})$ is a twice differentiable function on $T^*\mathbb{T}^n \times \mathbb{R}$ and satisfies the following:

(1) $\sup \left\|\nabla^2 H_t\right\| < K$

(2) $C\|\boldsymbol{v}\|^2 < \langle H_{\boldsymbol{pp}}(\boldsymbol{z}, t)\boldsymbol{v}, \boldsymbol{v}\rangle < C^{-1}\|\boldsymbol{v}\|^2$ for some positive C independant of $(\boldsymbol{z}, t)$ and $\boldsymbol{v} \neq 0$.

Sometimes Hamiltonian systems such that $H_{\boldsymbol{pp}}$ is definite positive are called *optical*. This is why we refer to Assumption 1 as one of *uniform opticity*.

Assumption 2 (Assymptotic quadraticity)

$H(\boldsymbol{q}, \boldsymbol{p}, t)$ is a C^2 function on $T^*\mathbb{T}^n$ satisfying the following:

(1) $\det H_{\boldsymbol{pp}} \neq 0$.

(2) For $\|\boldsymbol{p}\| \geq K_1$, $\quad H(\boldsymbol{q}, \boldsymbol{p}, t) = \langle A\boldsymbol{p}, \boldsymbol{p}\rangle + \boldsymbol{c}.\boldsymbol{p}, \qquad A^t = A, \det A \neq 0.$

Here A denotes a constant matrix, and $\boldsymbol{c}$ a constant in $\mathbb{R}^n$. We stress that A (and hence $H_{\boldsymbol{pp}}$) is not necessarily positive definite.

Theorem 39.5 *Let h^ϵ be the time ϵ of a Hamiltonian flow for a Hamiltonian function satisfying any of the Assumptions 1 or 2. Then, for small enough ϵ, h^ϵ is a symplectic twist map of $T^*\mathbb{T}^n$ (or on U, respectively).*

Remark 39.6 Proposition 39.5 holds for $h_t^{t+\epsilon}$ whenever it does for h^ϵ: $h_t^{t+\epsilon}$ is the time ϵ of the Hamiltonian $G(\boldsymbol{z}, s) = H(\boldsymbol{z}, t+s)$, which satisfies all the assumptions H does.

Proof. We prove the proposition with Assumption 1, and indicate how to adapt the proof to the other assumption. We can work in the covering space $\mathbb{R}2n$ of $T^*\mathbb{T}^n$, to which the flow lifts. The differential of h^t at a point $\boldsymbol{z} = (\boldsymbol{q}, \boldsymbol{p})$ is solution of the linear *variational equation* [9]

[9] In general, if ϕ^t is solution of the O.D.E. $\dot{\boldsymbol{z}} = X_t(\boldsymbol{z})$ then $D\phi^t$ is solution of $\dot{U}(t) = DX_t(\phi^t \boldsymbol{z})U(t), U(0) = Id$. Heuristically, this can be seen by differentiating $\frac{d}{dt}\phi^t(\boldsymbol{z}) = X_t(\phi^t(\boldsymbol{z}))$ with respect to $\boldsymbol{z}$ (see e.g. Hirsh & Smale (1974)).

(39.2) $$\dot{U}(t) = J\nabla^2 H(h^t(\boldsymbol{z}))U(t), \quad U(0) = Id, \quad J = \begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix}$$

We first prove that $U(\epsilon)$ is not too far from Id:

Lemma 39.7 *Consider the linear equation:*

$$\dot{U}(t) = A(t)U(t), \quad U(t_0) = U_0$$

where U and A are $n \times n$ matrices and $\|A(t)\| < K, \forall t$. Then :

$$\|U(t) - U_0\| < K\,\|U_0\|\,|t - t_0|e^{K|t-t_0|}.$$

Proof. Let $V(t) = U(t) - U_0$, so that $V(t_0) = 0$. We have:

$$\begin{aligned}\dot{V}(t) &= A(t)\left(U(t) - U_0\right) + A(t)U_0 \\ &= A(t)V(t) + A(t)U_0\end{aligned}$$

and hence:

$$\|V(t)\| = \|V(t) - V(t_0)\| \leq \int_{t_0}^{t} K\,\|V(s)\|\,ds + |t - t_0|K\,\|U_0\|$$

For all $|t - t_0| \leq \epsilon$, we can apply Gronwall's inequality (see Hirsh & Smale (1974)) to get:

$$\|V(t)\| \leq \epsilon K\,\|U_0\|\,e^{K|t-t_0|}$$

and we get the result by setting $\epsilon = |t - t_0|$. □

We now proceed with the proof of Proposition 39.5. By Lemma 39.7 we can write:

$$U(\epsilon) - Id = \int_0^{\epsilon} J\nabla^2 H(h^s(\boldsymbol{z})).(Id + O_1(s))ds$$

where $\|O_1(s)\| < 2Ks$, for ϵ, and hence s, small enough.

Let $(\boldsymbol{q}(t), \boldsymbol{p}(t)) = h^t(\boldsymbol{q}, \boldsymbol{p}) = h^t(\boldsymbol{z})$. The matrix $\boldsymbol{b}_\epsilon(\boldsymbol{z}) = \partial\boldsymbol{q}(\epsilon)/\partial\boldsymbol{p}$, is the upper right $n \times n$ matrix of $U(\epsilon)$. It is given by:

(39.3) $$\boldsymbol{b}_\epsilon(\boldsymbol{z}) = \int_0^{\epsilon} H_{\boldsymbol{pp}}(h^s(\boldsymbol{z}))ds + \int_0^{\epsilon} O_2(s)ds$$

where $\left|\int_0^{\epsilon} O_2(s)ds\right| < K^2\epsilon^2$. From this, and the fact that

(39.4) $$C\,\|\boldsymbol{v}\|^2 < \langle H_{\boldsymbol{pp}}(\boldsymbol{z})\boldsymbol{v}, \boldsymbol{v}\rangle < C^{-1}\,\|\boldsymbol{v}\|^2,$$

we deduce that:

(39.5) $$(\epsilon C - K^2\epsilon^2)\,\|\boldsymbol{v}\|^2 < \langle \boldsymbol{b}_\epsilon(\boldsymbol{z})\boldsymbol{v}, \boldsymbol{v}\rangle < (\epsilon C^{-1} + K^2\epsilon^2)\,\|\boldsymbol{v}\|^2$$

so that in particular $\boldsymbol{b}_\epsilon(\boldsymbol{z})$ is nondegenerate for small enough ϵ. Since $\boldsymbol{b}_\epsilon(\boldsymbol{z})$ is periodic in $\boldsymbol{z}$, the set of nonsingular matrices $\{\boldsymbol{b}_\epsilon(\boldsymbol{z})\}_{\boldsymbol{z}\in\mathbb{R}2n}$ is included in a compact set and thus:

(39.6) $$\sup_{z\in \mathbb{R}2n} \left\| \boldsymbol{b}_\epsilon^{-1}(\boldsymbol{z}) \right\| < K',$$

for some positive K'. We can now apply Proposition 26.4to show that h^ϵ is a symplectic twist map with a generating function S defined on all of $\mathbb{R}2n$.

Remark 39.8 The above proof shows that h^ϵ satisfies a certain *convexity condition* which can be useful in finding minimal orbits (see Chapter MIN):

(39.7) $$\langle \boldsymbol{b}_\epsilon^{-1}\boldsymbol{v}, \boldsymbol{v}\rangle = \left\langle \left(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}(\epsilon)\right)^{-1} \boldsymbol{v}, \boldsymbol{v} \right\rangle \geq a\left\|\boldsymbol{v}\right\|^2, \quad \forall \boldsymbol{v} \in \mathbb{R}^n.$$

where a is a positive constant. To see that it is the case, note that, denoting by

$$m = \inf_{\|\boldsymbol{v}\|=1,\ \boldsymbol{z}\in\mathbb{R}2n} \left\|\boldsymbol{b}_\epsilon^{-1}(\boldsymbol{z})\right\|$$

and M the corresponding sup, (39.5) implies:

$$m(\epsilon C - K^2\epsilon^2)\|\boldsymbol{v}\|^2 < \langle \boldsymbol{b}_\epsilon^{-1}(\boldsymbol{z})\boldsymbol{v}, \boldsymbol{v}\rangle < M(\epsilon C^{-1} + K^2\epsilon^2)\left\|\boldsymbol{v}\right\|^2.$$

We now adapt the above proof to Assumption 2. Note that under this assumption, we can still derive (39.3) : the boundary condition (2) implies that $\nabla^2 H$ is bounded. Since H is C^2, and $H_{\boldsymbol{pp}} = A$ outside a compact set, $H_{\boldsymbol{pp}}(h^s\boldsymbol{z})$ is uniformly close to $H_{\boldsymbol{pp}}(\boldsymbol{z})$ for small s, and thus the first matrix integral in (39.3) is non singular for $\boldsymbol{z}$ and small s. Thus $\boldsymbol{b}_\epsilon(\boldsymbol{z})$ is also nonsingular for small ϵ. Since $\boldsymbol{b}_\epsilon(\boldsymbol{z}) = \epsilon A$ outside of the compact set $\|\boldsymbol{p}\| \leq K_1$, the set of matrices $\{\boldsymbol{b}_\epsilon(\boldsymbol{z}) \mid |\boldsymbol{z} \in \mathbb{R}^n\}$ is compact and hence (39.5) holds, which proves the proposition in this case. □

C. Decomposition Of Hamiltonian Maps Into Twist Maps

When, as is the case in Theorems 39.2and 39.5, the time ϵ maps of a Hamiltonian system are all symplectic twist maps , one can readily decompose the time 1 map into such twist maps. Take a time dependent Hamiltonian, for example. Its time 1 map h^1 can be written:

$$h^1 = (h^{\frac{1}{N}})^N$$

and, for N large enough, each $h^{\frac{1}{N}}$ is a symplectic twist map . It is only slightly more complicated when H is time dependent. In this case we can write:

(39.8) $$h^1 = h^1_{\frac{N-1}{N}} \circ (h^{\frac{N-1}{N}}_{\frac{N-2}{N}}) \circ \ldots h^{\frac{k+1}{N}}_{\frac{k}{N}} \circ \ldots h_0^{\frac{1}{N}}$$

and each $h^{\frac{k+1}{N}}_{\frac{k}{N}}$ is an symplectic twist map by assumption on our Hamiltonian. as the next Proposition shows.

What may be more surprising, and gives strength to this method, is that there is a large class of Hamiltonian systems which, even though their time ϵ is *not* twist, can be decomposed into a product of symplectic twist maps. This is a generalization of an idea that LeCalvez (astérisque) applied in his variational proof of the Poincaré-Birkhoff Theorem.

This will work with either of the following, very broad, assumptions:

Assumption 3.

H is a C^2 function on $T^*M \times [0,1]$, and the domain U is a compact neighborhood in T^*M.

Assumption 4.

$H(\boldsymbol{z},t) = H_t(\boldsymbol{z})$ is a function on $T^*\mathbb{T}^n \times \mathbb{R}$ satisfying $\sup \|\nabla^2 H_t\| < K$.

Proposition 39.9 (Decomposition) *Let $H(\boldsymbol{z},t)$ be a Hamiltonian function satisfying Asssumptions 3 or 4, or the hypothesis of either Theorem 39.2 or Theorem 39.5. Then the time 1 h^1 of its corresponding Hamiltonian system can be decomposed into a finite product of symplectic twist maps:*

$$h^1 = F_{2N} \circ \ldots \circ F_1.$$

Proof. We have given the trivial proof above for Hamiltonians that satisfies the hypothesis of Theorems 39.2 and 39.5. We now prove the proposition when H satisfies Assumption 3. Pick a ball bundle $U(K) = \{(\boldsymbol{q},\boldsymbol{p}) \mid \|\boldsymbol{p}\| \leq K\}$ with K large enough so that $U \subset U(K)$. Let G be the time s of the geodesic flow, where s is chosen so that G is an symplectic twist map on $U(K)$. That such an s exists is proven in Proposition 38.2 . We can write:

$$\begin{aligned} h^1 &= G \circ \left(G^{-1} \circ h^1_{\frac{N-1}{N}}\right) \circ G \circ \ldots \circ \left(G^{-1} \circ h^{\frac{k+1}{N}}_{\frac{k}{N}}\right) \circ \ldots \circ G \circ \left(G^{-1} \circ h_0^{\frac{1}{N}}\right) \\ &= F_{2N} \circ \ldots \circ F_1. \end{aligned} \tag{39.9}$$

One can check that, at each successive step of the decomposition, the points remain in $U(K)$. Our new G is an symplectic twist map , by assumption, and $G^{-1} \circ h^{\frac{k+1}{N}}_{\frac{k}{N}}$ is an symplectic twist map by openess of the set of twist maps on a compact neighborhood (see Exercise STMstmopen).

Suppose now that H satisfies Assumption 4. Let $G(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}+\boldsymbol{p},\boldsymbol{p})$, our favorite symplectic twist map (see, *eg.* Example STMstandardexample) on $T^*\mathbb{T}^n$. Decompose h^1 as in Equation (39.9) . We now show that $G^{-1} \circ h^{\frac{k+1}{N}}_{\frac{k}{N}}$ is also a symplectic twist map. Lemma 39.4 implies that $h_t^{t+\epsilon}$ satisfies $\|Dh_t^{t+\epsilon} - Id\| < \epsilon K e^{K\epsilon}$. Hence

$$\left\|DG^{-1}.Dh^{\frac{k+1}{N}}_{\frac{k}{N}} - DG^{-1}\right\| < C\frac{1}{N}e^{\frac{K}{N}}$$

for some positive constant C. Thus $G^{-1} \circ h^{\frac{k+1}{N}}_{\frac{k}{N}}$ is twist for N large enough, since the sufficient conditions $\det\, \partial \boldsymbol{Q}/\partial \boldsymbol{p} \neq 0$ and $\|(\partial Q/\partial p)^{-1}\| < \infty$ are both open with respect to the C^1 norm.

□

40. SUSPENSION OF SYMPLECTIC TWIST MAPS BY HAMILTONIAN FLOWS

Moser (1986) showed how to suspend a monotone twist map of the annulus into a time 1 map of a (time dependant) Hamiltonian system satisfying the fiber convexity $H_{pp} > 0$. In subsection A we present a suspension theorem for higher dimensional symplectic twist maps announced by M. Bialy and L. Polterovitch, which implies Moser's theorem in two dimensions. These authors kindly agreed to let their complete proof appear for the first time in this book. In subsection B, we give the proof, due to the author, of a suspension theorem where we let go of a symmetry condition assumed by Bialy and Polterovitch. The price we pay is the loss of the fiber convexity of the suspending Hamiltonian.

A. SUSPENSION WITH FIBER CONVEXITY

Theorem 40.1 (Bialy and Polterovitch) *Let F be a symplectic twist map with generating function S satisfying:*

(40.1) $\partial_{12}S(\boldsymbol{q}, \boldsymbol{Q})$ *is symmetric and negative nondegenerate.*

Then there exists a smooth Hamiltonian function $H(\boldsymbol{q}, \boldsymbol{p}, t)$ on $T^*\mathbb{T}^n \times [0,1]$ convex in the fiber (i.e. $H_{\boldsymbol{pp}}$ is positive definite) such that F is the time 1 map of the Hamiltonian flow generated by H. The Hamiltonian function H can also be made periodic in the time t.

Proof. Following Moser, we will construct a Lagrangian function $L(\boldsymbol{q}, \boldsymbol{v}, t)$ on $\mathbb{R}2n \times [0,1]$ with the following properties:

(40.2) (a) The corresponding solutions of the Euler-Lagrange equations connecting the points $\boldsymbol{q}$ and $\boldsymbol{Q}$ in the covering space $\mathbb{R}^n$ in the time interval $[0,1]$ are straight lines $\boldsymbol{q} + t(\boldsymbol{Q} - \boldsymbol{q})$;

(40.2) (b) $S(\boldsymbol{q}, \boldsymbol{Q}) = \int_0^1 L(\boldsymbol{q} + t(\boldsymbol{Q} - \boldsymbol{q}), \boldsymbol{Q} - \boldsymbol{q}, t)dt$;

(40.2) (c) L is strictly convex with respect to $\boldsymbol{v}$: $\frac{\partial^2 L}{\partial \boldsymbol{v}^2}$ is positive definite.

(40.2) (d) $L(\boldsymbol{q} + \boldsymbol{m}, \boldsymbol{v}, t) = L(\boldsymbol{q}, \boldsymbol{v}, t)$ for all $\boldsymbol{m}$ in $\mathbb{Z}^n$.

If such a function L is constructed, its Legendre transform H satisfies the conclusion of Theorem 40.1: (40.2) (a) and (b) imply that F is the time 1 map of the Hamiltonian H, (40.2) (c) implies that $H_{\boldsymbol{pp}}$ is convex (see Exercise 47.2) and (40.2) (d) that the Euler-Lagrange flow of L takes place on $T\mathbb{T}^n$ and hence the Hamiltonian flow of H is defined on $T^*\mathbb{T}^n$.

Note that if (40.2) (c) is satisfied then (40.2) (a) is equivalent to the following equation:

(40.2) (a') $\dfrac{\partial^2 L}{\partial \boldsymbol{v} \partial \boldsymbol{q}} \boldsymbol{v} + \dfrac{\partial^2 L}{\partial \boldsymbol{v} \partial t} - \dfrac{\partial F}{\partial \boldsymbol{q}} = 0.$

Lemma 40.2 *Set* $R_{ij}(\boldsymbol{q}, \boldsymbol{v}, t) = -\dfrac{\partial^2 S}{\partial \boldsymbol{q}_i \partial Q_j}(\boldsymbol{q} - t\boldsymbol{v}, \boldsymbol{q} + (1-t)\boldsymbol{v})$. *Then the following holds:*

(40.3) (a) $R_{ij} = R_{ji}$;

(40.3) (b) $\dfrac{\partial R_{ij}}{\partial v_k} = \dfrac{\partial R_{ik}}{\partial v_j}$;

(40.3) (c) $\dfrac{\partial R_{ij}}{\partial q_k} = \dfrac{\partial R_{ik}}{\partial q_j}$;

(40.3) (d) $\dfrac{\partial R_{ij}}{\partial t} + \sum_l \dfrac{\partial R_{lj}}{\partial q_i} v_l = 0$

for all i, j, k.

The proof is straightforward and uses the fact that the matrix $\frac{\partial^2 S}{\partial \boldsymbol{q} \partial \boldsymbol{Q}}$ is symmetric.

Lemma 40.3 *Set* $L(\boldsymbol{q}, \boldsymbol{v}, t) = \int_0^1 (1-\lambda) \sum_{i,j} R_{ij}(\boldsymbol{q}, \lambda \boldsymbol{v}, t) v_i v_j d\lambda$. *Then the following holds:*

(40.4) (a) $\dfrac{\partial L}{\partial v_i} = \displaystyle\int_0^1 \sum_j R_{ij}(\boldsymbol{q}, \tau \boldsymbol{v}, t) d\tau$

(40.4) (b) $\dfrac{\partial^2 L}{\partial v_i \partial v_j} = R_{ij}$

(40.4) (c) *L satisfies Equation (40.2) (a').*

Proof. Rewrite L as follows:

$$\begin{aligned} L(\boldsymbol{q}, \boldsymbol{v}, t) &= \int_0^1 \int_\lambda^1 ds \sum_{i,j} R_{ij}(\boldsymbol{q}, \lambda \boldsymbol{v}, t) v_i v_j d\lambda = \int_0^1 ds \int_0^s d\lambda \sum_{i,j} R_{ij}(\boldsymbol{q}, \lambda \boldsymbol{v}, t) v_i v_j \\ &= \int_0^1 ds \int_0^1 s \sum_{i,j} R_{ij}(\boldsymbol{q}, s\tau \boldsymbol{v}, t) v_i v_j d\tau = \int_0^1 \sum_i v_i \alpha_i(\boldsymbol{q}, s\boldsymbol{v}, t) ds, \end{aligned} \tag{40.5}$$

where $\alpha_i(\boldsymbol{q}, \boldsymbol{v}, y) = \displaystyle\int_0^1 \sum_j R_{ij}(\boldsymbol{q}, \tau \boldsymbol{v}, t) v_j d\tau$. We can rewrite the last integral of (40.5) as a path integral:

$$\int_0^1 \sum_i v_i \alpha_i(\boldsymbol{q}, s\boldsymbol{v}, t) ds = \int_\gamma \sum_i \alpha_i dv_i,$$

where $\gamma(s) = (\boldsymbol{q}, s\boldsymbol{v}, t)$. Fixin $\boldsymbol{q}$ and t, Equation (40.3) (b) implies that the form $\sum_i \alpha_i dv_i$ is closed, and, because $\boldsymbol{v} \in \mathbb{R}^n$, exact, say $\sum_i \alpha_i dv_i = dA$ for some function $A(\boldsymbol{v})$ on $\mathbb{R}^n$. Then the Fundamental Theorem of Calculus yields:

$$L(\boldsymbol{q}, \boldsymbol{v}, t) = A(\boldsymbol{v}) - A(0).$$

Since $\sum_i \alpha_i dv_i = dA = \frac{\partial L}{\partial \boldsymbol{v}} d\boldsymbol{v}$, Equation (40.4) (a) follows. The proof of (40.4) (b) is similar. We now prove (40.4) (c). In view of (40.4) (a), the left hand side I of (40.2) (a)' can be written as follows:

$$\begin{aligned} I =& \sum_l v_l \int_0^1 \sum_j \frac{\partial r_{ij}}{\partial q_l}(\boldsymbol{q}, \tau \boldsymbol{v}, t) v_j d\tau + \int_0^1 \sum_j \frac{\partial R_{ij}}{\partial t}(\boldsymbol{q}, \tau \boldsymbol{v}, t) v_j d\tau \\ &+ \int_0^1 (1-\lambda) \sum_{i,j} \frac{\partial R_{lj}}{\partial q_i}(\boldsymbol{q}, \lambda \boldsymbol{v}, t) v_i v_j d\lambda. \\ =& a_1 + a_2 - a_3, \end{aligned}$$

where a_k is the k^{th} integral in the above expression. Rewrite a_3 using (40.3) (c) as follows:

$$a_3 = \int_0^1 \sum_{l,j} \frac{\partial R_{ij}}{\partial q_j} v_l v_j d\tau - \int_0^1 \sum_{l,j} \frac{\partial R_{l,j}}{\partial q_i} v_l v_j \tau d\tau.$$

The first term is equal to a_1. Therefore:

$$I = \int_0^1 \sum_j v_j \left\{ \frac{\partial R_{ij}}{\partial t}(\boldsymbol{q}, \tau \boldsymbol{v}, t) + \sum_{l,j} \frac{\partial R_{l,j}}{\partial q_i} \tau v_l \right\} d\tau.$$

Equation (40.3) implies that the bracket, and hence I, vanish. □

Given any function $L(\boldsymbol{q}, \boldsymbol{v}, t)$, set

$$\tilde{L}(\boldsymbol{q}, \boldsymbol{Q}) = \int_0^1 L(\boldsymbol{q} + t(\boldsymbol{Q} - \boldsymbol{q}), \boldsymbol{Q} - \boldsymbol{q}, t) dt.$$

Lemma 40.4 *Assume that L satisfies (40.2) (a'). Then the following holds:*

$$(40.6)\ (a) \qquad \frac{\partial \tilde{L}}{\partial q_i} = -\frac{\partial L}{\partial v_i}(\boldsymbol{q}, \boldsymbol{Q}-\boldsymbol{q}, 0);$$

$$(40.6)\ (b) \qquad \frac{\partial \tilde{L}}{\partial Q_i} = \frac{\partial L}{\partial v_i}(\boldsymbol{Q}, \boldsymbol{Q}-\boldsymbol{q}, 1);$$

$$(40.6)\ (c) \qquad \frac{\partial^2 \tilde{L}}{\partial q_i \partial Q_j} = -\frac{\partial^2 L}{\partial v_i \partial v_j}(\boldsymbol{q}, \boldsymbol{Q}-\boldsymbol{q}, 0).$$

Proof. Equation (40.6) (c) is a consequence of (40.6) (a), which we now prove. The same argument also proves (40.6) (b). It is not hard to check that if L satisfies (40.2) (a)' then:

$$\frac{d}{dt}\left\{\frac{\partial L}{\partial v_i}(\boldsymbol{q}+t(\boldsymbol{Q}-\boldsymbol{q}), \boldsymbol{Q}-\boldsymbol{q}, t)\right\} = \frac{\partial L}{\partial q_i}(\boldsymbol{q}+t(\boldsymbol{Q}-\boldsymbol{q}), \boldsymbol{Q}-\boldsymbol{q}, t).$$

Therefore,

$$\frac{\partial \tilde{L}}{\partial q_i}(\boldsymbol{q}, \boldsymbol{Q}) =$$
$$\int_0^1 \left\{-\frac{\partial L}{\partial v_i}(\boldsymbol{q}+t(\boldsymbol{Q}-\boldsymbol{q}), \boldsymbol{Q}-\boldsymbol{q}, t) + (1-t)\frac{d}{dt}\left(\frac{\partial L}{\partial v_i}(\boldsymbol{q}+t(\boldsymbol{Q}-\boldsymbol{q}), \boldsymbol{Q}-\boldsymbol{q}, t)\right)\right\} dt$$
$$= \int_0^1 \frac{d}{dt}\left\{(1-t)\frac{\partial L}{\partial v_i}(\boldsymbol{q}+t(\boldsymbol{Q}-\boldsymbol{q}), \boldsymbol{Q}-\boldsymbol{q}, t)\right\} dt = -\frac{\partial L}{\partial vi}(\boldsymbol{q}, \boldsymbol{Q}-\boldsymbol{q}, 0).$$

□

Given any two differentiable functions $L(\boldsymbol{q}, \boldsymbol{v}, t), f(\boldsymbol{q}, t)$, set:

$$L_f(\boldsymbol{q}, \boldsymbol{v}, t) = F(\boldsymbol{q}, \boldsymbol{v}, t) + \frac{\partial f}{\partial q}(\boldsymbol{q}, t)\boldsymbol{v} + \frac{\partial f}{\partial t}(\boldsymbol{q}, t).$$

Lemma 40.5

(40.7) (a) $\tilde{L}_f(\boldsymbol{q}, \boldsymbol{Q}) = \tilde{L}(\boldsymbol{q}, \boldsymbol{Q}) + f(\boldsymbol{Q}, 1) - f(\boldsymbol{q}, 0)$;

(40.7) (b) *If L satisfies* (40.2) (a') *then L_f satisfies it as well, for all f.*

The proof of this lemma is straightforward. We are now in position to finish the proof of Theorem 40.1. Let L be the function defined in Lemma 40.3. From (40.6) (c) and (40.4) (b), we get:

$$\frac{\partial^2 \tilde{L}}{\partial q_i \partial \boldsymbol{Q}_j}(\boldsymbol{q}, \boldsymbol{Q}) = -\frac{\partial^2 \tilde{L}}{\partial v_i \partial v_j}(\boldsymbol{q}, \boldsymbol{Q}-\boldsymbol{q}, 0) = \frac{\partial^2 S}{\partial q_i \partial Q_j}(\boldsymbol{q}, \boldsymbol{Q}),$$

and therefore

$$\tilde{L}(\boldsymbol{q}, \boldsymbol{Q}) = S(\boldsymbol{q}, \boldsymbol{Q}) + A(\boldsymbol{q}) + b(\boldsymbol{Q})$$

for some differentiable functions a and b. Set

$$f(\boldsymbol{q}, t) = (1-t)A(\boldsymbol{q}) - tb(\boldsymbol{Q}).$$

We claim that the function L_f satisfies (40.2) (a)-(d). We prove these properties one by one.

1. We proved in (40.4) (c) that L satisfies (40.2) (a'), and hence (40.2) (a). Equation (40.7) (b) proves that L_f does as well.

2. From (40.7) (a), we get:

$$\tilde{L}_f(\boldsymbol{q}, \boldsymbol{Q}) = \tilde{L}(\boldsymbol{q}, \boldsymbol{Q}) - b(\boldsymbol{Q}) - A(\boldsymbol{q}) = S(\boldsymbol{q}, \boldsymbol{Q}),$$

which proves (40.2) (b).

3. $\dfrac{\partial^2 \tilde{L}_f}{\partial \boldsymbol{v}^2} = \dfrac{\partial^2 \tilde{L}}{\partial \boldsymbol{v}^2} = (R_{ij}) = -\dfrac{\partial^2 S}{\partial \boldsymbol{q} \partial \boldsymbol{Q}}(\boldsymbol{q} - t\boldsymbol{v}, \boldsymbol{q} + (1-t)\boldsymbol{v})$. Since this last matrix is positive definite by (40.1) , so is the first one.

4. Since $S(\boldsymbol{q}+\boldsymbol{m}, \boldsymbol{Q}+\boldsymbol{m}) = S(\boldsymbol{q}, \boldsymbol{Q})$, the function L is periodic in $\boldsymbol{q}$. We need to check that $\frac{\partial f}{\partial t}$ and $\frac{\partial f}{\partial q}$ are also periodic in $\boldsymbol{q}$. Using the definitions and (40.6) (a) and (b), one can easily check that

$$\tilde{L}(\boldsymbol{q}, \boldsymbol{q}) = \frac{\partial \tilde{L}}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{q}) = \frac{\partial \tilde{L}}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{q}) = 0.$$

From the definitions of the functions a and b we obtain that

$$A(\boldsymbol{q}) + b(\boldsymbol{q}) = -S(\boldsymbol{q}, \boldsymbol{q}), \quad \frac{\partial a}{\partial \boldsymbol{q}} = -\frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{q}), \quad \frac{\partial b}{\partial \boldsymbol{q}}(\boldsymbol{q}) = -\frac{\partial S}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{q}).$$

Because of the periodicity of S, all these functions are periodic in $\boldsymbol{q}$. Since

$$\frac{\partial f}{\partial t} = (1-t)\frac{\partial a}{\partial \boldsymbol{q}} - t\frac{\partial b}{\partial \boldsymbol{q}}, \quad \frac{\partial f}{\partial \boldsymbol{q}} = -a - b,$$

both $\frac{\partial f}{\partial t}$ and $\frac{\partial f}{\partial q}$ are periodic. This finishes the proof of our claim, and hence that of Theorem 40.1 . □

B. SUSPENSION WITHOUT CONVEXITY

If we let go of the symmetry of $\frac{\partial^2 S}{\partial q \partial Q}$ (but keep some form of definiteness) in Theorem 40.1 , we can still suspend the twist map F by a Hamiltonian flow. The cost is relatively high however: we can no longer insure that the Hamiltonian is convex in the fiber. The proof, quite different from that of Theorem 40.1 , first appeared in Golé (1994c) .

Theorem 40.6 *Let $F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})$ be a symplectic twist map of $T^*\mathbb{T}^n$ whose differential $\boldsymbol{b}(\boldsymbol{z}) = \frac{\partial \boldsymbol{Q}(\boldsymbol{z})}{\partial \boldsymbol{p}}$ satisfies:*

$$\text{(40.8)} \qquad \sup_{\boldsymbol{z} \in T^*\mathbb{T}^n} \langle \boldsymbol{b}^{-1}(\boldsymbol{z})\boldsymbol{v}, \boldsymbol{v} \rangle > a \left\| \boldsymbol{v} \right\|, \quad a > 0, \ \forall \boldsymbol{v} \neq 0 \in \mathbb{R}^n.$$

Then F is the time 1 map of a (time dependant) Hamiltonian H.

Remark 40.7 Condition (40.8) tells us that F does not twist infinitely much.

Proof. Let $S(\boldsymbol{q}, \boldsymbol{Q})$ be the generating function of F. Since $\boldsymbol{p} = -\partial_1 S(\boldsymbol{q}, \boldsymbol{Q})$, we have that $\boldsymbol{b} = \partial Q / \partial \boldsymbol{p} = -\left(\partial_{12} S(\boldsymbol{q}, \boldsymbol{Q})\right)^{-1}$. Hence equation (40.8) translates into:

$$\text{(40.9)} \qquad \sup_{(\boldsymbol{q}, \boldsymbol{Q}) \in \mathbb{R}2n} \langle -\partial_{12} S(\boldsymbol{q}, \boldsymbol{Q})\boldsymbol{v}, \boldsymbol{v} \rangle > a \left\| \boldsymbol{v} \right\|, \quad a > 0, \forall \boldsymbol{v} \neq 0 \in \mathbb{R}^n.$$

The following lemma show that (40.9) implies the hypothesis of Proposition 26.4, which in turn shows that whenever we have a function on $\mathbb{R}2n$ which is suitably periodic and satisfies (40.9) , it is the generating function for some symplectic twist map.

Lemma 40.7 *Let $\{A_x\}_{x\in\Lambda}$ be a family of $n\times n$ real matrices satisfying:*

$$\sup_{x\in\Lambda}|\langle A_x\boldsymbol{v},\boldsymbol{v}\rangle| > a\|\boldsymbol{v}\|^2,\quad \forall \boldsymbol{v}\neq 0\in\mathbb{R}^n.$$

Then :

$$\sup_{x\in\Lambda}\left\|A_x^{-1}\right\| < a^{-1}.$$

We postpone the proof of this lemma to the end. We now construct a differentiable family $S_t,\ t\in[0,1]$ of generating functions, with $S_1=S$, and then show how to make a Hamiltonian vector field out of it, whose time 1 map is F. Let

$$S_t(\boldsymbol{q},\boldsymbol{Q}) = \begin{cases} \frac{1}{2}af(t)\|\boldsymbol{Q}-\boldsymbol{q}\|^2 & \text{for } 0<t\le\frac{1}{2} \\ \frac{1}{2}af(t)\|\boldsymbol{Q}-\boldsymbol{q}\|^2 + (1-f(t))S(\boldsymbol{q},\boldsymbol{Q}) & \text{for } \frac{1}{2}\le t\le 1. \end{cases}$$

where f is a smooth positive functions, $f(1)=f'(1/2)=0,\ f(1/2)=1$ and $\lim_{t\to 0^+}f(t)=+\infty$. We will ask also that $1/f(t)$, which can be extended continuously to $1/f(0)=0$, be differentiable at 0. The choice of f has been made so that S_t is differentiable with respect to t, for $t\in(0,1]$. Furthermore, it is easy to verify that:

$$\sup_{(\boldsymbol{q},\boldsymbol{Q})\in\mathbb{R}2n}\langle -\partial_{12}S_t(\boldsymbol{q},\boldsymbol{Q})\boldsymbol{v},\boldsymbol{v}\rangle > a\left\|\boldsymbol{v}\right\|^2,\quad a>0, \forall\boldsymbol{v}\neq 0\in\mathbb{R}^n, t\in(0,1].$$

Hence S_t generates a smooth family $F_t,\ t\in(0,1]$ of symplectic twist maps, and in fact $F_t(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}+(af(t))^{-1}\boldsymbol{p},\boldsymbol{p}),\quad t\le 1/2)$, so that $\lim_{t\to 0^+}F_t = Id$, in any topology that one desires (on compact sets). Let us write

$$s_t(\boldsymbol{q},\boldsymbol{p}) = S_t\circ\psi_t(\boldsymbol{q},\boldsymbol{p}),$$

where ψ_t is the change of coordinates given by the fact that F_t is twist. It is not hard to verify that $\psi_t(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q},\boldsymbol{q}-(af(t))^{-1}\boldsymbol{p}),\quad t\le 1/2$. so that:

$$s_t(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}(af(t))^{-2}\|\boldsymbol{p}\|^2$$

In particular, by our assumption on $1/f(t)$, s_t can be differentiably continued for all $t\in[0,1]$, with $S_0\equiv 0$. Hence, in the $\boldsymbol{q},\boldsymbol{p}$ coordinates, we can write:

$$F_t^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = ds_t,\quad t\in[0,1].$$

By Theorem 47.7, F_t is a Hamiltonian isotopy. □

Proof of Lemma 40.7

For all non zero $\boldsymbol{v}\in\mathbb{R}^n$, we have:

$$\inf_{x\in\Lambda}\frac{|\langle A_x\boldsymbol{v},\boldsymbol{v}\rangle|}{\|\boldsymbol{v}\|^2} > a$$

But:

$$\inf_{\boldsymbol{v}\in\mathbb{R}^n}\frac{|\langle A_x\boldsymbol{v},\boldsymbol{v}\rangle|}{\|\boldsymbol{v}\|^2} = \inf_{\|\boldsymbol{v}\|=1}|\langle A_x\boldsymbol{v},\boldsymbol{v}\rangle| \le \inf_{\|\boldsymbol{v}\|=1}\|A_x\boldsymbol{v}\|$$

so that $\inf_{x\in\Lambda}\inf_{\|\boldsymbol{v}\|=1}\|A_x\boldsymbol{v}\| > a$. But:

$$\inf_{\|\boldsymbol{v}\|=1}\|A_x\boldsymbol{v}\| = \inf_{\boldsymbol{v}\in\mathbb{R}^n-\{0\}}\frac{\|A_x\boldsymbol{v}\|}{\|\boldsymbol{v}\|} = inf_{\boldsymbol{v}\in\mathbb{R}^n-\{0\}}\frac{\|\boldsymbol{v}\|}{\left\|A_x^{-1}\boldsymbol{v}\right\|}$$

so that, finally:

$$\sup_{x\in\Lambda}\left\|A_x^{-1}\right\| = \left(\inf_{x\in\Lambda}\inf_{\boldsymbol{v}\in\mathbb{R}^n-\{0\}}\frac{\|\boldsymbol{v}\|}{\|A_x\boldsymbol{v}\|}\right)^{-1} < a^{-1}.$$

□

41.1 Return Maps In Hamiltonian Systems

A. RETURN MAPS OF HAMILTONIAN SYSTEMS ARE SYMPLECTIC

Consider a time independent Hamiltonian on $\mathbb{R}^{2n+2}$, with its standard symplectic structure $\Omega_0 = \sum_{k=0}^n dq_k \wedge dp_k$. Assume that we have a periodic trajectory γ for the Hamiltonian flow. It must then lie in the energy level $H = H(\gamma(0))$, since H is time independent. Take any $2n+1$ dimensional open disk $\tilde{\Sigma}$ which is transverse to γ at $\gamma(0)$, and such that $\tilde{\Sigma}$ intersects γ only at $\gamma(0)$.

Fig. 41. 2.

Such a disk clearly always exists, if γ is not a fixed point. In fact, one can assume that, in a local Darboux chart, $\tilde{\Sigma}$ is the hyperplane with equation $q_0 = 0$: this is because in the construction of Darboux coordinates, one can start by choosing an arbitrary nonsingular differentiable function as one of the coordinate function (see Arnold (1978), section 43, or Weinstein (1979) , Extension Theorem, lecture 5.)

Define $\Sigma = \tilde{\Sigma} \cap \{H = H_0\}$. It is a standard fact (true for periodic orbits of *any* C^1 flow) that the Hamiltonian flow h^t admits a Poincaré return map $\mathcal{R}$, defined on Σ around $\boldsymbol{z}_0$, by $\mathcal{R}(\boldsymbol{z}) = h^{t(\boldsymbol{z})}(\boldsymbol{z})$, where $t(\boldsymbol{z})$ is the first return time of $\boldsymbol{z}$ to Σ under the flow (see Hirsh & Smale (1974), Chapter 13).

We claim that $\mathcal{R}$ is symplectic, with the symplectic structure induced by Ω_0 on Σ.

Since $\tilde{\Sigma}$ is transverse to γ, we may assume that:

$$\dot{q}_0 = \frac{\partial H}{\partial p_0} \neq 0$$

on $\tilde{\Sigma}$. Hence, by the Implicit Function Theorem, the equation

$$H(0, q_1 \ldots, q_n, p_0, \ldots, p_n) = H_0$$

implies that p_0 is a function of $(q_1, \ldots, q_n, p_1, \ldots, p_n)$. This makes the latter variables a system of local coordinates for Σ, and since $dq_0 = 0$ on Σ, the restriction of Ω_0 is in fact

$$\omega = \Omega_0\big|_{\Sigma} = \sum_{k=1}^{n} dq_k \wedge dp_k.$$

To prove that $\mathcal{R}$ is symplectic, remember that, by (41.-1) , for any closed curve in Σ, or more generally for any closed 1–chain c in Σ,

$$\int_{\mathcal{R}c} \boldsymbol{p}d\boldsymbol{q} - Hdt = \int_c \boldsymbol{p}d\boldsymbol{q} - Hdt$$

since c and $\mathcal{R}c$ are on the same trajectory tube. Here $\mathcal{R}c$ represent the chain in $\mathbb{R}^{2n+2} \times \mathbb{R}$ given by $(\mathcal{R}(c(s)), t^{c(s)})$.

This equality implies that the function $S(\boldsymbol{z}) = \int_{\boldsymbol{z}_0}^{\boldsymbol{z}} \mathcal{R}^*(\boldsymbol{p}d\boldsymbol{q} - Hdt) - (\boldsymbol{p}d\boldsymbol{q} - Hdt)$ is well defined. But, on Σ, the differential of the form inside this integral is $\mathcal{R}^*\omega - \omega$, since both dq_0 and dH are zero there. Hence $\mathcal{R}^*\omega - \omega = d^2S = 0$, i.e., $\mathcal{R}$ is symplectic. □

B. TWISTING AROUND ELLIPTIC FIXED POINTS

We now follow Moser (1977). If 0 is an elliptic fixed point, that is $DR(0)$ has all its eigenvalues on the unit circle, a normal form theorem ???(find ref.) says that (generically?) the map R is, around 0 given by:

$$\begin{aligned} Q_k &= q_k cos\Phi_k(\boldsymbol{q},\boldsymbol{p}) - p_k sin\Phi_k(\boldsymbol{q},\boldsymbol{p}) + f_k(\boldsymbol{q},\boldsymbol{p}) \\ P_k &= q_k sin\Phi_k(\boldsymbol{q},\boldsymbol{p}) + p_k cos\Phi_k(\boldsymbol{q},\boldsymbol{p}) + g_k(\boldsymbol{q},\boldsymbol{p}) \\ \Phi_k(\boldsymbol{q},\boldsymbol{p}) &= \alpha_k + \sum_{l=1}^{n} \beta_{kl}(q_l^2 + p_l^2). \end{aligned}$$

where the error term f_k, g_k are C^1 and have vanishing derivatives up to order 3 at the origin. We now show how this map is, in "polar coordinates" a symplectic twist map of $T^*\mathbb{T}^n$, *whenever the matrix* $\{\beta_{kl}\}$ *is non singular*. Let V be a punctured neighborhood of 0 such that: $0 < \sum_k (q_k^2 + p_k^2) < \epsilon$. We introduce on V new coordinates (r_k, θ_k) by:

$$\boldsymbol{q}_k = \sqrt{2r_k\epsilon}\cos 2\pi\theta_k' \quad p_k = \sqrt{2r_k\epsilon}\sin 2\pi\theta_k$$

where θ_k is determined modulo 1. One can check that V is transformed into the "annular" set:

$$U = \{(r_k, \theta_k) \in \mathbb{T}^n \times \mathbb{R}^n \mid \sum_k \left(2r_k - \frac{1}{2n}\right)^2 < \frac{1}{4n^2}\}$$

Since the symplectic form $dq \wedge dp$ is transformed into $\epsilon d\boldsymbol{r} \wedge d\boldsymbol{\theta}$, R remains symplectic in these new coordinates, with the symplectic form $d\boldsymbol{r} \wedge d\boldsymbol{\theta}$. In fact, it is exact symplectic in U. Remember that to check this, it is enough to show that, for any closed curve γ:

$$\int_{R\gamma} \boldsymbol{r}d\boldsymbol{\theta} = \int_\gamma \boldsymbol{r}d\boldsymbol{\theta}.$$

It is easy to see that $2\epsilon r_k d\theta_k = p_k dq_k - q_k dp_k$, so by Stokes' theorem:

$$2\epsilon \int_\gamma \boldsymbol{r}d\boldsymbol{\theta} = \int_{\partial D} \boldsymbol{p}d\boldsymbol{q} - \boldsymbol{q}d\boldsymbol{p} = -2\int_D \omega$$

where D is a 2 manifold in V with boundary $\partial D = \gamma$. Since R preserves ω in V, it must preserve the last integral, and hence the first. To see that R satisfies the two other conditions for being a symplectic twist map, we just write it in the new coordinates:

$$\begin{aligned}\Theta_k &= \theta_k + \psi_k(\boldsymbol{r}) + o_1(\epsilon)\\ R_k &= r_k + o_1(\epsilon)\\ \psi_k &= \alpha_k + \epsilon \sum_{l+1}^{n} 2\beta_{kl} r_l .\end{aligned}$$

where $\epsilon^{-1} o_1(\epsilon, \boldsymbol{\theta}, \boldsymbol{r})$ and its first derivatives in $\boldsymbol{r}, \boldsymbol{\theta}$ tend to 0 uniformally as $\epsilon \to 0$. We can rewrite this as:

$$\mathcal{R}(\boldsymbol{\theta}, \boldsymbol{r}) = (\boldsymbol{\theta} + \epsilon \boldsymbol{B}\boldsymbol{r} + \boldsymbol{\alpha} + o_1(\epsilon), \boldsymbol{r} + o_1(\epsilon)) .$$

So for small ϵ, the condition $\det \partial\boldsymbol{\Theta}/\partial\boldsymbol{r} \neq 0$ is given by the nondegeneracy of $\boldsymbol{B} = \{\beta_{kl}\}$, one uses the fact that $\mathcal{R}$ is C^1 close to a completely integrable symplectic twist map to show that $\mathcal{R}$ is twist in U (the twist condition is open.) The fact that it is homotopic to Id derives from Exercise 23.2.

Note that the set V and therefore U are not invariant under $\mathcal{R}$. However, it is still possible to show the existence of infinitely many periodic points for $\mathcal{R}$: this is the content of the Birkhoff– Lewis theorem (see Moser (1977)) .

Remarks HAMrem and HAMgrad are 39.8 , Corollary HAMpartial is 38.4 , Proposition HAMdecompone and HAMdecomptwo are (39.8) and 39.9 , Theorem HAMexp is 38.1 , HAMhamstm is 39.2 . Section HAMsecgeom is 38.0, Theorem HAMthmbp is 40.1

CHAPTER 8 or HAMP

PERIODIC ORBITS FOR HAMILTONIAN SYSTEMS

March 14 1999

More about my recent results in the case $\mathbb{T}^n$ (sphere linking)? Define the radius of injectivity. Put HAMpartial here instead of in HAM? Add the proof of Theorem 33.5? Copy the precise statement of Arnold's conjecture.

We present here some results of existence and multiplicity of periodic orbits in Hamiltonian systems on cotangent bundles. Our main goal is to show the power, and relative simplicity of the method of decomposition by symplectic twist map as presented in Chapter HAM, which results into finite dimensional variational problems. Some of the results in this chapter have recently been improved upon by other authors. However, this was done at a high price, using hard analytic and topological method. Many of these, and other improvements could probably be obtained through the method presented here.

Bla bla bla....

42. Periodic Orbits In The Cotangent Of The n-Torus

We present here two results of existence and multiplicity of periodic orbits for Hamiltonian systems in $T^*\mathbb{T}^n$. The are easy corollaries of the Theorems of existence of multiple periodic orbits for symplectic twist maps proven in Chapter PSTM. The first one concerns a certain class of optical systems, the second one Hamiltonians that are quadratic nondegenerate outside of a bounded set.

A. Optical Hamiltonians

Assumption 42.1 (Uniform Opticity) $H(\boldsymbol{q}, \boldsymbol{p}, t) = H_t(\boldsymbol{z})$ is a twice differentiable function on $T^*\mathbb{T}^n \times \mathbb{R}$ (or $T^*M \times \mathbb{R}$, where $\tilde{M} = \mathbb{R}^n$) and satisfies the following:

(1) $\sup \left\|\nabla^2 H_t\right\| < K$

(2) The matrices $H_{\boldsymbol{pp}}(\boldsymbol{z}, t)$ are positive definite and its smallest eigenvalue are uniformly bounded below by $C > 0$.

Theorem 42.2 *Let $H(\boldsymbol{q}, \boldsymbol{p}, t)$ be a Hamiltonian function on $T^*\mathbb{T}^n \times \mathbb{R}$ satisfying Assumption 42.1. Then the time 1 map h^1 of the associated Hamiltonian flow has at least $n+1$ periodic orbits of type $\boldsymbol{m}, d$, for each prime $\boldsymbol{m}, d$, and 2^n when they are all non degenerate.*

Proof. We can decompose the time 1 map:

$$h^1 = h^1_{\frac{N-1}{N}} \circ \ldots \circ h^{\frac{k+1}{N}}_{\frac{k}{N}} \circ \ldots \circ h_0^{\frac{1}{N}}.$$

and each of the maps $h^{\frac{k+1}{N}}_{\frac{k}{N}}$ is the time $\frac{1}{N}$ of the (extended) flow, starting at time $\frac{k}{N}$. Proposition HAMde-componeof Chapter 4 shows that, for N big enough, such maps are symplectic twist maps. Moreover, we also noted in Chapter HAM, Remark HAMremthat these maps also satisfy the convexity condition . The result follows from Theorem STMPthesis. □

B. ASYMPTOTICALLY QUADRATIC HAMILTONIANS

We now turn to systems that are not necessarily optical, but satisfy a certain quadratic "boundary condition" which makes them completely integrable outside a compact set:

Theorem 42.3 *Let $H : T^*\mathbb{T}^n \times \mathbb{R} \to \mathbb{R}$ satisfy the following boundary condition:*

$$H(\boldsymbol{q}, \boldsymbol{p}, t) = \frac{1}{2}\langle A\boldsymbol{p}, \boldsymbol{p}\rangle + \boldsymbol{c}\cdot\boldsymbol{p}, \quad A^t = A, \det A \neq 0 \text{ when } \|\boldsymbol{p}\| \geq K. \tag{42.1}$$

Then h^1, the time–1 map of the Hamiltonian flow has at least $n+1$ distinct $\boldsymbol{m}, d$–orbits, and 2^n when they are all nondegenerate (i.e. generically). Furthermore, such an orbit lays entirely in the set $\|\boldsymbol{p}\| \leq K$ if and only if the rotation vector $\boldsymbol{m}/d$ belongs to the ellipsoid:

$$\mathcal{E} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \left\|A^{-1}(\boldsymbol{x} - \boldsymbol{c})\right\| \leq K\}.$$

Proof. The boundary condition (42.1) is Assumption 2 preceeding Theorem 39.5 , in which it is proven that the time ϵ of such Hamiltonians are twist maps. Hence, as remarked in Proposition HAMdecomptwo, the time 1 map can be decomposed into symplectic twist maps. To insure that these twist maps satisfy the conditions of Theorem STMPtquad, we go back to the proof of that proposition, and note that, instead of $G(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}+\boldsymbol{p}, \boldsymbol{p})$, we can take $G(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}+A\boldsymbol{p}+\boldsymbol{c}, \boldsymbol{p})$, the time 1 map of $H_0(\boldsymbol{q}, p) = \frac{1}{2}\langle A\boldsymbol{p}, \boldsymbol{p}\rangle + \boldsymbol{c}.\boldsymbol{p}$, obviously a symplectic twist map . Then, outside the set $\|\boldsymbol{p}\| \leq K$, the maps F_{2k}, F_{2k-1} of the decomposition are respectively the time 1 and the time $(\frac{1}{N} - 1)$ of the Hamiltonian flow associated to H_0, that is:

$$F_{2k}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + A\boldsymbol{p} + \boldsymbol{c}, \boldsymbol{p})$$
$$F_{2k-1}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + \frac{1-N}{N}(A\boldsymbol{p} + \boldsymbol{c}), \boldsymbol{p}).$$

These maps clearly satisfy the conditions of Theorem STMPtquad, which proves the existence of the advertised number of $\boldsymbol{m}, d$ orbits.

To localize these orbits, note that an orbit starting in $\|\boldsymbol{p}\| \geq K$ must stay there, and the map h^1 on such an orbit is just G. The rotation number of such an orbit is thus

$$(\boldsymbol{Q} - \boldsymbol{q}) = A\boldsymbol{p} + \boldsymbol{c}$$

from which we conclude that $\boldsymbol{m}/d$ is in the complement of $\mathcal{E}$. □

Remark 42.4 There is a distinction between periodic orbits of h^1 and periodic orbits of the Hamiltonian equations: for a general time dependent Hamiltonian flow, $(h^1)^n \neq h^n$, and hence an $\boldsymbol{m}, d$ periodic orbit for h^1 is not necessarily one for the O.D.E. (which should satisfy $h^{t+d}(\boldsymbol{z}) = h^t(\boldsymbol{z}) + (k\boldsymbol{m}, 0)$ for all $t \in [kd, (k+1)d), k \in \mathbb{Z}$). However, if H is periodic in time, of period 1, the equality $(h^1)^n = h^n$ does hold, and in this case the two notions coincide. In particular, this holds trivially for time independant Hamiltonians. Unfortunatly, these cases are degenerate in our setting, since $Dh^d(\boldsymbol{z})$ preserves the vector field X_H, which is thus an eigenvector with eigenvalue one. So in these cases, we can only claim the cuplength estimates for the number of periodic orbits for the Hamiltonian *flow* in either Theorem 42.2 or 42.3. We think that some further argument should yield, even in the time periodic case the sum of the betti number estimate for the number of flow periodic orbits, when the periodic orbits are nondegenerate *as orbits of the flow*: i.e., the only eigenvector of eigenvalue one for $Dh^d(\boldsymbol{z})$ is in the direction of the vector field X_H.

C. BIBLIOGRAPHY...

43. Periodic Orbits In General Cotangent Spaces

We now turn to the study of Hamiltonian systems in cotangent spaces of arbitrary compact manifolds. Our main result, which first appeared in Golé (1994) is:

Theorem 43.1 *Let (M, g) be a compact Riemannian manifold. Let $F : T^*M \to T^*M$ be the time 1 map of a time dependent Hamiltonian H on B^*M, where H is a C^2 function satisfying the boundary condition:*

$$H(\boldsymbol{q}, \boldsymbol{p}, t) = g(\boldsymbol{q})(\boldsymbol{p}, \boldsymbol{p}) \text{ for } \|\boldsymbol{p}\| \geq C.$$

where C is strictly smaller than the radius of injectivity ???have to define it somewhere???. Then F has $cl(M)$ distinct fixed points and $sb(M)$ if they are all non degenerate. Moreover, these fixed points lie inside the set $\{\|\boldsymbol{p}\| < C\}$ and can all be chosen to correspond to homotopically trivial closed orbits of the Hamiltonian flow.

THE DISCRETE VARIATIONAL SETTING

Define

$$B^*M = \{(\boldsymbol{q}, \boldsymbol{p}) \in T^*M \mid g(\boldsymbol{q})(\boldsymbol{p}, \boldsymbol{p}) = \|\boldsymbol{p}\|^2 \leq C^2 < R^2\},$$

where R is the radius of injectivity of (M, g). Let π denote the canonical projection $\pi : B^*M \to M$. Let F be as in Theorem 43.1. From Proposition HAMdecomptwoin Chapter HAM, we can decompose F into a product of symplectic twist maps :

$$F = F_{2N} \circ \ldots \circ F_1,$$

where F_{2k} restrained to the boundary ∂B^*M of B^*M is the time 1 map h_0^1 of the geodesic flow with Hamiltonian $H_0(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{p}\|^2$. Likewise, F_{2k-1} is $h_0^{\frac{1-N}{N}}$ on ∂B^*M.

Let S_k be the generating function for the twist map F_k and $\psi_k = \psi_{F_k}$ the diffeomorphism $(q, p) \to (q, Q)$ induced by the twist condition on F_k. We can assume that ψ_k is defined on a neighborhood U of B^*M in T^*M. Let

$$O = \{\overline{\boldsymbol{q}} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{2N}) \in M^{2N} \mid (\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \in \psi_k(U) \text{ and } (\boldsymbol{q}_{2N}, \boldsymbol{q}_1) \in \psi_{2N}(U)\} \tag{44.1}$$

O is an open set in M^{2N}, containing a copy of M (the elements $\overline{\boldsymbol{q}}$ such that $\boldsymbol{q}_k = \boldsymbol{q}_1$, for all k).

Next, define :

$$W(\overline{\boldsymbol{q}}) = \sum_{k=0}^{2N} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}), \tag{4.2}$$

where we have set $\boldsymbol{q}_{2N+1} = \boldsymbol{q}_1$. Choosing to work in some local coordinates around $\overline{\boldsymbol{q}} \in M^{2N}$, we let $\boldsymbol{p}_k = -\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ and $\boldsymbol{P}_k = \partial_2 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$. In other words, $(\boldsymbol{q}_k, \boldsymbol{p}_k) \in T^*_{\boldsymbol{q}_k} M$ is such that $\psi_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ and $(\boldsymbol{q}_{k+1}, \boldsymbol{P}_k) \in T^*_{\boldsymbol{q}_{k+1}} M$ is such that $F_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{P}_k)$. We let the reader check that the following proofs can be written in coordinate free notation.

As in the case $M = \mathbb{T}^n$, we have:

Lemma 44.1 (Critical Action Principle) *The sequence $\overline{\boldsymbol{q}}$ of O is a critical point of W if and only if the sequence $\{(\boldsymbol{q}_k, \boldsymbol{p}_k)\}_{k \in \{1, \ldots, 2N, 1\}}$ is an orbit under the successive F_k's, that is if and only if $(\boldsymbol{q}_1, \boldsymbol{p}_1)$ is a fixed point for F.*

Proof. Because the twist maps are exact symplectic and using the definitions of $\boldsymbol{p}_k$, $\boldsymbol{P}_k$, we have:

$$\boldsymbol{P}_k d\boldsymbol{q}_{k+1} - \boldsymbol{p}_k d\boldsymbol{q}_k = dS_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}), \tag{44.2}$$

and hence

$$dW(\overline{\boldsymbol{q}}) = \sum_{k=1}^{2N} (\boldsymbol{P}_{k-1} - \boldsymbol{p}_k) d\boldsymbol{q}_k$$

which is null exactly when $\boldsymbol{P}_{k-1} = \boldsymbol{p}_k$, i.e. when $F_k(\boldsymbol{q}_{k-1}, \boldsymbol{p}_{k-1}) = (\boldsymbol{q}_k, \boldsymbol{p}_k)$. Now remember that we assumed that $\boldsymbol{q}_{2N+1} = \boldsymbol{q}_1$. □

Hence, to prove Theorem 43.1, we need to find enough critical points for W. As before, we will study the gradient flow of W (where the gradient will be given in terms of the metric g) and use the boundary condition to find an isolating block. The main difference with the previous situations on $T^*\mathbb{T}^n$ is that we cannot put W in the general framework of generating phases quadratic at infinity. Nonetheless, thanks to the boundary condition we imposed on the Hamiltonian, we are able to construct an isolating block and use Floer's theorem of continuation to get a grasp on the topology of the invariant set, and hence on the number of critical points.

45. Proof Of Theorem 43.1

THE ISOLATING BLOCK

In this subsection we prove that the set B defined as follows:

$$B = \{\overline{\boldsymbol{q}} \in O \mid \|\boldsymbol{p}_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})\| \le C\} \tag{45.1}$$

is an isolating block for the gradient flow of W, where O is defined in (44.1) , C is as in the hypotheses of Theorem 43.1 and $\boldsymbol{p}_k = -\partial_1 S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$. Note that when $\big\|\boldsymbol{p}_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})\big\| = C$,

$$\mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = |a_k|C \quad \text{where} \begin{cases} a_k = 1 & \text{if } k \text{ is even} \\ a_k = \frac{1-N}{N} & \text{if } k \text{ is odd} \end{cases} \tag{45.2}$$

Clearly B contains the constant sequences, a set homeomorphic to M.

Proposition 45.1 *B is an isolating block for the gradient flow of W.*

Proof. Suppose that the point $\overline{\boldsymbol{q}}$ of U is in the boundary of B. This means that $\|\boldsymbol{p}_k\| = C$ for at least one k. As noted in (45.2) , this means that $\mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) = |a_k|C$ for some factor a_k only depending on the parity of k. We want to show that this distance increases either in positive or negative time along the gradient flow of W. This flow is given by:

$$\dot{\boldsymbol{q}}_k = A_k(\boldsymbol{P}_{k-1} - \boldsymbol{p}_k) = \nabla W_k(\overline{\boldsymbol{q}}) \tag{45.3}$$

where $A_k = A(\boldsymbol{q}_k)$ is the inverse of the matrix of coefficients of the metric g at the point $\boldsymbol{q}_k$. Remember that we have put the product metric on O, induced by its inclusion in M^{2N} (see Remark HAMgradon the definition of the gradient of a function).

We compute the derivative of the distance along the gradient flow at a boundary point of B, using Corollary HAMpartialand the fact that $h_0^{a_k}(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{P}_k)$:

$$\begin{aligned} \frac{d}{dt}\mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})\big|_{t=0} &= \partial_1 \mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \cdot \nabla W_k(\overline{\boldsymbol{q}}) \\ &\quad + \partial_2 \mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \cdot \nabla W_{k+1}(\overline{\boldsymbol{q}}) \\ &= sign(a_k)\frac{-\boldsymbol{p}_k}{\|\boldsymbol{p}_k\|} \cdot A_k(\boldsymbol{P}_{k-1} - \boldsymbol{p}_k) \\ &\quad + sign(a_k)\frac{\boldsymbol{P}_k}{\|\boldsymbol{P}_k\|} \cdot A_{k+1}(\boldsymbol{P}_k - \boldsymbol{p}_{k+1}) \end{aligned} \tag{45.4}$$

We now need a simple linear algebra lemma to treat this equation.

Lemma 45.1 *Let $\langle\, ,\, \rangle$ denote a positive definite bilinear form in $\mathbb{R}^n$, and $\|.\|$ its corresponding norm. Suppose that $\mathbf{p}$ and $\mathbf{p}'$ are in $\mathbb{R}^n$,that $\|\mathbf{p}\| = C$ and that $\|\mathbf{p}'\| \le C$. Then :*

$$\langle \mathbf{p} \ , \ \mathbf{p}' - \mathbf{p} \ \rangle \le 0.$$

Moreover, equality occurs if and only if $\mathbf{p}' = \mathbf{p}$.

Proof. From the positive definiteness of the metric, we get:

$$\langle \ \mathbf{p}' - \mathbf{p}, \mathbf{p}' - \mathbf{p} \ \rangle \ge 0,$$

with equality occuring if and only if $\mathbf{p}' = \mathbf{p}$. From this, we get:

$$2\langle \ \mathbf{p}, \mathbf{p}' \ \rangle \le \langle \ \mathbf{p}', \mathbf{p}' \ \rangle + \langle \ \mathbf{p}, \mathbf{p} \ \rangle$$

Finally,

$$\langle \ (\mathbf{p}' - \mathbf{p}), \mathbf{p} \ \rangle = \langle \ \mathbf{p}', \mathbf{p} \ \rangle - \langle \ \mathbf{p}, \mathbf{p} \ \rangle \le \frac{1}{2}(\langle \ \mathbf{p}', \mathbf{p}' \ \rangle - \langle \ \mathbf{p}, \mathbf{p} \ \rangle) \le 0$$

with equality occuring if and only if $\mathbf{p}' = \mathbf{p}$. □

Applying Lemma 45.1 to each of the right hand side terms in (45.4) , we can deduce that $\frac{d}{dt}\mathrm{Dis}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ is positive when k is pair, negative when k is odd. Indeed, because of the boundary condition in the hypothesis of the theorem, we have $\|\boldsymbol{P}_k\| = \|\boldsymbol{p}_k\|$ whenever $\|\boldsymbol{p}_k\| = C$: the boundary ∂B^*M is invariant under F and all the F_k's. On the other hand $\overline{\boldsymbol{q}} \in B \Rightarrow \|\boldsymbol{p}_l\| \leq C$ and $\|\boldsymbol{P}_l\| \leq C$, for all l, by invariance of B^*M. Finally, a_k is positive when k is even, negative when k is odd.

We have shown that the gradient flow exits B at all the points of ∂B except perhaps at the edges of ∂B. These edges are the sets of points $\overline{\boldsymbol{q}}$ such that more than one $\boldsymbol{p}_k$ has norm C. The problem at these edges occurs when k is in an interval $\{l, \ldots, m\}$ such that, for all j in this interval, $\left\|\boldsymbol{p}_j\right\| = C = \|\boldsymbol{P}_j\|$ and $\nabla W_j(\overline{\boldsymbol{q}}) = 0$.

It is now crucial to note that $\{l, \ldots, m\}$ can not cover all of $\{0, \ldots, 2N\}$: this would mean that $\overline{\boldsymbol{q}}$ is a critical point corresponding to a fixed point of h_0^1 in ∂B^*M. But such a fixed point is forbidden by our choice of C: orbits of our Hamiltonian on the set $\|\boldsymbol{p}\| = C$ are geodesics, but geodesics in that energy level can not be fixed loops since $C > 0$, and they can not close up in time one either since C is less than the injectivity radius.

We now let $k = m$ in (45.4) and see that the flow must definitely escape the set B at $\overline{\boldsymbol{q}}$ in either positive or negative time, from the m^{th} face of B. □

Remark 45.2 If the Hamiltonian considered is optical and we decompose its time 1 map into a product of N twist maps as in HAMdecompone, all the F_k's coincide with $h_0^{\frac{1}{N}}$ on the boundary of B^*M. In that case, all the a_k's in the above proof are positive, and B is a repeller block in this case.

END OF PROOF OF THEOREM 43.1

To finish the proof of Theorem 43.1 we use Floer's theorem TOPOfloerthmof continuation of normally hyperbolic invariant sets. We consider the family F_λ of time 1 maps of the Hamiltonians:

$$H_\lambda = (1 - \lambda)H_0 + \lambda H.$$

Corresponding to this is a family of gradient flows ζ_λ^t, solution of

$$\frac{d}{dt}\overline{\boldsymbol{q}} = \nabla W_\lambda(\overline{\boldsymbol{q}}),$$

where W_λ is the discrete action corresponding to the decomposition in symplectic twist maps of the map F_λ. We take care that this decomposition has the same number of steps $2N$ for each λ. The manifold on which we consider these (local) flows is O, which is an open neighborhood of B in M^{2N}. Each of the F_λ satisfies the hypothesis of Theorem 43.1, and thus Proposition 45.1 applies to ζ_λ^t for all λ in $[0, 1]$: B is an isolating block for each one of these flows. Hence the maximum invariant sets G_λ for the flows ζ_λ^t in B are related by continuation. The part of Floer's Theorem that we need to check is that G_0 is a normally hyperbolic invariant manifold for ζ_0^t.

Lemma 45.2 *Let $G_0 = \{\overline{\boldsymbol{q}} \in B \mid \boldsymbol{q}_k = \boldsymbol{q}_1, \forall k\}$. Then G_0 is a normally hyperbolic invariant set for ζ_0^t. G_0 is a retract of O and it is the maximal invariant set in B.*

Proof. The only critical points for W_0 in B are the points of G_0 which correspond to restpoints of the geodesic flow, *i.e.* the zero section. Indeed, critical points of W_0 in B corresponds to periodic points of period 1 for the geodesic flow in B^*M. Our definition of that sets precludes nontrivial periodic geodesics in B^*M. We now show thatthe maximum invariant set for ζ_0^t in B is included in G_0. Since ζ_0^t is a gradient flow, such an invariant set is formed by critical points and connecting orbits between them. The only critical points of W_0 in B are the points of G_0. If there were a connecting orbit entirely in B, it would have to connect two points in G_0, which is absurd since $W_0 \equiv 0$ on G_0, whereas W_0 should increase along non constant orbits.

G_0 is a retract of M^{2N} under the composition of the maps:

$$\overline{\boldsymbol{q}} = (\boldsymbol{q}_1, \dots, \boldsymbol{q}_{2N}) \to \boldsymbol{q}_1 \to (\boldsymbol{q}_1, \boldsymbol{q}_1, \dots, \boldsymbol{q}_1) = \alpha(\overline{\boldsymbol{q}})$$

which is obviously continuous and fixes the points of G_0.

It remains to show that G_0 is normally hyperbolic. Since $G_0 \cong M$ is an n-dimensional manifold made of critical points, saying that it is normally hyperbolic is equivalent to saying that $ker \nabla^2 W_0(\overline{\boldsymbol{q}})$ has dimension n: indeed, if it is the case, the only possible vectors in this kernel must be tangent to G_0, and thus he differential of the flow is nondegenerate on the normal space to TG_0. In the present situation, the second variation formula of Lemma 31.2says that the 1-eigenspace of Dh_0^1 is isomorphic to the kernel of $\nabla^2 W_0$. Hence it is enough to check that at a point $(\boldsymbol{q}_1, 0) \in B^*M$ corresponding to $\overline{\boldsymbol{q}}$, 1 is an eigenvalue of multiplicity exactly n for $Dh_0^1(\boldsymbol{q}_1, 0)$. Let us compute $Dh_0^1(\boldsymbol{q}_1, 0)$ in local coordinates. It is the solution at time 1 of the linearized (or variation) equation:

$$\dot{U} = J\nabla^2 H_0(\boldsymbol{q}_1, 0)U$$

along the constant solution $(q(t), p(t)) = (\boldsymbol{q}_1, 0)$, where J denotes the usual symplectic matrix $\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$. An operator solution for the above equation is given by $exp\left(tJ\nabla^2 H_0(\boldsymbol{q}_1, 0)\right)$ On the other hand:

$$\nabla^2 H_0(\boldsymbol{q}_1, 0) = \begin{pmatrix} 0 & 0 \\ 0 & A(\boldsymbol{q}_1) \end{pmatrix}$$

which we computed from $H_0(q, p) = A(q)p.p$, the zero terms appearing at $p = 0$ because they are either quadratic or linear in p. From this,

$$Dh_0^1(\boldsymbol{q}_1, 0) = exp\left(J\nabla^2 H_0(\boldsymbol{q}_1, 0)\right) = \begin{pmatrix} I & A(\boldsymbol{q}_1) \\ 0 & I \end{pmatrix}$$

is easily derived. This matrix has exactly n independent eigenvectors of eigenvalue 1 (it has in fact no other eigenvector). Hence, from Lemma 31.2, $\nabla^2 W(\overline{\boldsymbol{q}})$ has exactly n vectors with eigenvalue 0, as was to be shown. □

We now conclude the proof of Theorem 43.1. We have proved that the gradient flow ζ^t, has an invariant set G_1 with $H^*(M) \hookrightarrow H^*(G_1)$. From this we get in particular:

$$cl(G_1) \geq cl(M) \text{ and } sb(G_1) \geq sb(M).$$

Theorem 50.2 tells us that ζ^t must have at least $cl(G_1)$ rest points in the set G_1, and $sb(G_1)$ if all rest points are nondegenerate. But Lemma 31.2tells us that nondegeneracy for $\nabla^2 W$ at a critical point is the same thing as nondegeneracy of a fixed point for F (no eigenvector of eigenvalue 1). This proves the existence of the advertised number of fixed points of the map F. In the following section, we will see that all these fixed points of the time 1 map correspond to periodic orbits may be chosen to be homotopically trivial. This concludes the proof of Theorem 43.1. □

FREE HOMOTOPY CLASSES

Since each F_k is close (or equal) to $h_0^{a_k}$ for some positive or negative a_k, we have: $\boldsymbol{q}$ is in the set $\psi_k(B^*_{\boldsymbol{q}}M)$ and, since $B^*_{\boldsymbol{q}}M \to \psi_k(B^*M)$ is a diffeomorphism, we can define a path $c_k(q, Q)$ between q and a point Q of $\psi_k(B^*_{\boldsymbol{q}}M)$ by taking the image by ψ_k of the oriented line segment between $\psi_k^{-1}(q)$ and $\psi_k^{-1}(Q)$ in $B^*_{\boldsymbol{q}}M$. In the case where $F_k = h_0^1$, this amounts to taking the unique geodesic between q and Q in $\psi_k(B^*_{\boldsymbol{q}}M)$.

If we look for periodic orbits of period d and of a given homotopy type, we decompose F^d into $2Nd$ twist maps, by decomposing F into $2N$. Analogously to (4.1), we define :

$$O_d = \{\overline{q} = (q_1, \dots, q_{2Nd}) \in M^{2Nd} \,|(q_k, q_{k+1}) \in \psi_k(U) \text{ and } (q_{2Nd}, q_1) \in \psi_{2Nd}(U)\},$$

remarking that the ψ_k's here correspond to the decomposition of F^d into $2Nd$ steps (U is as before a neighborhood of B^*M).

To each element $\overline{\boldsymbol{q}}$ in O_d, we can associate a closed curve $c(\overline{\boldsymbol{q}})$, made by joining up each pair (q_k, q_{k+1}) with the curve $c_k(q_k, q_{k+1})$ uniquely defined as above. This loop $c(\overline{\boldsymbol{q}})$ is piecewise differentiable and it depends continuously on $\overline{\boldsymbol{q}}$, and so do its derivatives (left and right). In the case of the decomposition of h_0^1 , taking F_k=h_0^1, this is exactly the construction of the broken geodesics (see Section 38.0). Now any closed curve in M belongs to a free homotopy class m.[10] To any d periodic point for F, we can associate a sequence $\overline{\boldsymbol{q}}(x) \in O_d$ of q coordinates of the orbit of this point under the successive F_k's in the decomposition of F^d.

Definition 45.3 Let $\boldsymbol{z}$ be a periodic point of period d for F. Let $\overline{\boldsymbol{q}}$ be the sequence in O_d corresponding to x. We say that x is an (m, d) point if $c(\overline{\boldsymbol{q}}(x))$ is in the free homotopy class m.

This definition has the advantage to make sense for any map F of T^*M which can be decomposed into the product of symplectic twist maps . If F is also the time 1 map of a Hamiltonian, it agrees with the obvious definition:

Proposition 45.4 *If $\boldsymbol{z}$ is an $\boldsymbol{m}, d$ periodic orbit, then the projection $\pi(\boldsymbol{z}(t)), t \in [0, d]$ of the orbit of $\boldsymbol{z}$ under the Hamiltonian flow is a closed curve in the free homotopy class $\boldsymbol{m}$.*

Proof. Left as an exercise (*Hint.* Use the geodesic flow to construct the homotopy between $c(\overline{\boldsymbol{q}}(\boldsymbol{z}))$ and $\pi(\boldsymbol{z}(t))$.)

Let

$$O_{m,d} = \{\overline{\boldsymbol{q}} \in O \mid c(\overline{\boldsymbol{q}}) \in m\} \tag{45.5}$$

Since $c(\overline{\boldsymbol{q}})$ depends continuously on $\overline{\boldsymbol{q}} \in O$, $O_{m,d}$ is a connected component of O. The reader who wants to make sure that, in the proof of Theorem 43.1, the orbits found are homotopically trivial, can check that the

[10] We remind the reader that free homotopy classes of loops differ from elements of $\pi_1(M)$ in that no base point is kept fixed under the homotopies. As a result, free homotopy classes can be seen as conjugacy classes in $\pi_1(M)$, and thus can not be endowed with a natural algebraic structure. Two elements of a free class give the same element in $H_1(M)$. Hence free homotopy classes form a set smaller than $\pi_1(M)$, bigger than $H_1(M)$. All these sets coincide if $\pi_1(M)$ is abelian.

proof we gave in last section works identically when one replace the space O, by its connected component $O_{e,1}$, where e is the homotopy class of the trivial curve. Another place where one uses this decomposition of O in different homotopy components is the following:

Theorem 45.5 *Let (M,g) be a Riemannian manifold of negative curvature and H be as in Theorem 1. If $\gamma_{\boldsymbol{m}}$ denotes the (unique) closed geodesic of free homotopy class $\boldsymbol{m}$, F has at least 2 (m,d) orbits in B^*M when $length(\gamma_{\boldsymbol{m}}) < dC$.*

The proof of Theorem 45.5 (see Golé (1994), Theorem 2) has the same broad outline as that of Theorem 1. We work in $O_{\boldsymbol{m},d}$ instead of O. The normally hyperbolic invariant set that we continue to in this setting is given by the set G_0 of critical sequences corresponding to the orbits under the $h_0^{a_k}$'s of the points on $\gamma_{\boldsymbol{m}}$. The normal hyperbolicity of G_0 derives this time from the hyperbolicity of the geodesic flow in negative curvature.

??? Add the proof in? ???

46. Linking Of Spheres: Toward A Generalization Of The Theorem Of Poincaré And Birkhoff

As stated in the introduction , Arnold conjectured in 1965 a generalization of the Theorem of Poincaré-Birkhoff for Hamiltonian maps of $\mathbb{T}^n \times \mathbb{B}^n$ (where $\mathbb{B}^n$ is the closed ball in $\mathbb{R}^n$).

Arnold's Linking of Spheres Conjecture

Generalized Arnold Conjecture Let M be a compact manifold, and F be a Hamiltonian map of a ball bundle B^*M in T^*M. Suppose that each sphere $\partial B^*_q M$ links with its image by F in ∂B^*M. Then F has at least $cl(M)$ distinct fixed points, and at least $sb(M)$ if they are nondegenerate.

In Banyaga & Golé (???)(see also Golé (1994)), we proved the simple case:

Theorem 46.1 *Let F be a symplectic twist map of B^*M which links spheres on the boundary ∂B^*M. Then F satisfies the generalized Arnold Conjecture.*

Proof. The proof of this theorem is trivial once one understands the meaning of the linking condition. If one looks at the Poincaré-Birkhoff situation, an easy equivalent condition to the boundary twist condition (points on the two boundary components go in opposite directions for some lift of F) is that a vertical fiber $\{x = x_0\}$ and its image by F should have a nonzero algebraic intersection number ($i.e.$ the number of intersections counted with orientation). Let us take this for the moment as a working definition of the linking of spheres in the general case:

Definition 46.2 (Boundary Twist: version 1) We say that a map $F : B^*M \to B^*M$ satisfies the boundary twist condition if each fiber $\Delta_{q_0} = \pi^{-1}(\boldsymbol{q}_0)$ intersects its image by F with a nonzero algebraic intersection number

We will see later on (for the reader who is comfortable with a little algebraic topology) that this intersection number condition is equivalent to linking of the boundary spheres as is usually defined in algebraic topology (and was probably meant by Arnold). The importance of this is that the boundary twist condition is indeed a topological condition on the action of the map *on the boundary.*

If F is a symplectic twist map, a fiber $\Delta_{\boldsymbol{q}}$ and its image under F may intersect at most once. Hence the boundary twist condition means in this case that all the fibers intersect their image *exactly once*. Fixed points of F correspond to critical points of $\boldsymbol{q} \to S(\boldsymbol{q}, \boldsymbol{q})$. This function is well defined since, by what preceeds, the diagonal in $M \times M$ is in the image of B^*M by the embedding ψ_F. Hence F has as many fixed points as the function $\boldsymbol{q} \to S(\boldsymbol{q}, \boldsymbol{q})$ has critical points on M. Morse and Lyusternick-Schnirelman's theories give the advertised estimates. □

We now show that, in the case considered by Arnold, our working definition of boundary twist is indeed equivalent to the classical one of algebraic topology. We first remind the reader of the classical definition of linking of spheres. Let $\Delta_{\boldsymbol{q}}$ be a fiber of B^*M as before. Then $\partial\Delta_{\boldsymbol{q}}$ is an n dimensional sphere. It make sense to talk about its linking with its image $F(\partial\Delta_{\boldsymbol{q}})$ in $\partial B^*\tilde{\mathbb{T}}^n$: the latter set has dimension $2n-1$ and the dimensions of the spheres add up to $2n-2$. The linking number $F(\partial\Delta_{\boldsymbol{q}})$ with $\partial\Delta_{\boldsymbol{q}}$ is given by the class $[F(\partial\Delta_{\boldsymbol{q}})] \in H_{n-1}(\partial B^*\tilde{\mathbb{T}}^n \backslash \partial\Delta_{\boldsymbol{q}})$ More precisely, we have:

$$\begin{aligned} H_{n-1}(\partial B^*\tilde{\mathbb{T}}^n \backslash \partial\Delta_{\boldsymbol{q}}) &\cong H_{n-1}\left(\mathbb{S}^{n-1} \times (\mathbb{R}^n - \{0\})\right) \\ &\overset{\text{Kunneth}}{\cong} H_{n-1}(\mathbb{S}^{n-1}) \oplus H_{n-1}(\mathbb{R}^n - \{0\}) \end{aligned} \qquad (46.1)$$

Thus, taking $\partial\Delta_{\boldsymbol{q}}$ from $\partial B^*\tilde{\mathbb{T}}^n$ creates a new generator in the $(n-1)$st homology, i.e. the generator b of $H_{n-1}(\mathbb{R}^n - \{0\})$.

The *linking number* of the spheres $F(\partial\Delta_{\boldsymbol{q}})$ and $\partial\Delta_{\boldsymbol{q}}$ is given by the $H_{n-1}(\mathbb{R}^n - \{0\}) \cong \mathbb{R}$ coefficient in the decomposition of the homology class $[F(\partial\Delta_{\boldsymbol{q}})]$ in the direct sum in (46.1) . If the linking number is nonzero, we say that the spheres $\partial\Delta_{\boldsymbol{q}}$ and its image by F *link*.

Definition (Boundary Twist: Version 2) We will say that the map F satisfies the *boundary twist condition* if for all $\boldsymbol{q} \in \tilde{\mathbb{T}}^n$ these spheres link in $\partial B^*\tilde{\mathbb{T}}^n$.

Lemma 46.3 *If F is the lift of a diffeomorphism of $B^*\mathbb{T}^n = \mathbb{T}^n \times B^n$, the two definitions of the boundary twist condition are equivalent. More precisely, the algebraic intersection number $\#(\Delta_{\boldsymbol{q}} \cap F(\Delta_{\boldsymbol{q}}))$ and the linking number of the spheres $\partial\Delta_{\boldsymbol{q}}$ and $F(\partial\Delta_{\boldsymbol{q}})$ are equal.*

Proof. We complete (46.1) into the following commutative diagram:

$$\begin{array}{ccc} H_{n-1}(\partial B^*\tilde{\mathbb{T}}^n \backslash \partial\Delta_{\boldsymbol{q}}) & \cong & H_{n-1}(\mathbb{R}^n - \{0\}) \oplus H_{n-1}(\mathbb{S}^n) \\ \downarrow i_* & & \downarrow j_* \\ H_{n-1}(B^*\tilde{\mathbb{T}}^n \backslash \Delta_{\boldsymbol{q}}) & \cong & H_{n-1}\left((\mathbb{R}^n - \{0\} \times B^n)\right) \end{array}$$

where i, j are inclusion maps. It is clear that $j_* b$ generates

$$H_{n-1}\left((\mathbb{R}^n - \{0\}) \times B^n\right) \cong H_{n-1}\left((\mathbb{R}^n - \{0\}) \times \mathbb{R}^n\right).$$

The last group measures the (usual) linking number of a sphere with the fiber Δ_q in $B^*\tilde{\mathbb{T}}^n \cong \mathbb{R}^{2n}$. But it is well known that such a number is the intersection number of any ball bounded by the sphere with the fiber Δ_q, counted with orientation. □

???more about my recent results in the case $\mathbb{T}^n$???

Theorem HAMPthmfp is 43.1, Theorem HAMPthmhyp is 45.5

CHAPTER 9 or AMG

*GENERALIZATIONS OF THE AUBRY-MATHER THEOREM

January 16, 2000

Complete

There are, strictly speaking, no *full* generalizations of the Aubry-Mather Theorem in higher dimensions: we will see in this chapter examples of fiber convex Lagrangian systems whose set of minimizers achieves only very few rotation directions. However some attempts of generalizations in higher dimensions are quite successful in what they try to achieve. In Section 47, we survey some results by de la Llave and his collaborators. Their setting is explicitly non dynamical but generalizes naturally the Frenkel-Kontorova model to functions on lattices of any dimension. They are entirely successful in proving an Aubry-Mather type theorem in this setting, as well as in some PDE cases. In Section 48, we review the work MacKay & Meiss (1992) who construct higher dimensional analogs of Aubry-Mather sets in symplectic twist maps that are close to the anti-integrable limit: one where the potential term in the generating function of a standard type map dominates. In Section 49, we survey the work of Mather on minimal measures in convex Lagrangian systems. This is the closest to a generalization of the Aubry-Mather theory as one can get in the setting of general convex Lagrangian systems (as well as symplectic twist maps). We start in Subsection A with an introduction to such minimizers and their relation to hyperbolic orbits. In Subsection B we give a quick review of some notions of ergodic theory that are needed in Subsection C, where we introduce minimal measures in Lagrangian systems. Subsection D explores, through examples, the intrinsic limitations of this theory. Section 50 shows that some of these limitations can be alleviated if one considers systems on cotangent bundles of hyperbolic manifolds.

47.* Aubry-Mather Theory for Functions on Lattices and PDE's.

A*. Functions on Lattices

Remember from Chapter 1 that the Frenkel–Kontorova model describes configurations of interacting particles in a periodic potential. For simplicity, these configurations are assumed to be one dimensional, and the interactions only involve nearest neighbors. The resulting energy function is the familiar:

$$W(\boldsymbol{x}) = \frac{1}{2}\sum_{k\in\mathbb{Z}}(x_k - x_{k+1})^2 - \sum_{k\in\mathbb{Z}} V(x_k)$$

where the potential function V has period 1. W coincides with the energy function for the standard map with generating function $S(x, X) = \frac{1}{2}(X - x)^2 - V(x)$. The variational equation $\nabla W = 0$ for this energy function is

$$(-\Delta \boldsymbol{x})_k - V'(x_k) = 0$$

where $\Delta(\boldsymbol{x})_k = 2x_k - x_{k-1} - x_{k+1}$ is the discretized Laplacian. Note that the configuration $\boldsymbol{x}$ can be seen as a function $\mathbb{Z} \to \mathbb{R}$ which to the integer k makes correspond the real x_k. One obtains (see Koch & al. (1994) , Candel & de la Llave (1997) , de la Llave (1999)) a natural generalization of this model, relevant to Statistical Mechanics, by asking that $\boldsymbol{x} : \mathbb{Z}^d \to \mathbb{R}$ be a function on a lattice of dimension d. We assume nearest neighbor interaction here. The energy becomes:

$$W(\boldsymbol{x}) = \frac{1}{2} \sum_{\{(k,j)\in\mathbb{Z}^d \mid |k-j|=1\}} (x_k - x_j)^2 - \sum_{k\in\mathbb{Z}^d} V(x_k).$$

Again V is of period 1 and the corresponding variational equation is still of the form:

$$(-\Delta \boldsymbol{x})_k - V'(x_k) = 0 \tag{47.1}$$

where $(\Delta \boldsymbol{x})_k = \sum_{|k-j|=1} x_j - 2dx_k$ is the d–dimensional discrete Laplacian. In fact, the theory in Candel & de la Llave (1997) applies to substantially more general settings, where k can belong to a set Λ on which a certain type of groups acts in a mildly prescribed way, and where the interactions involves not just nearest neighbors, but all possible pairs of particles (with some decay condition at infinity).

Remember that the solutions $\boldsymbol{x} : \mathbb{Z} \to \mathbb{R}$ found by Aubry and Mather for the Frenkel–Kontorova model are such that $|x_k - k\omega| \leq \infty$. One way to express this is by saying that the graph of $\boldsymbol{x} : \mathbb{Z} \to \mathbb{R}$ is at bounded distance from a line of slope ω in $\mathbb{R}\times\mathbb{R}$. Likewise, the following generalization of the Aubry-Mather Theorem finds configurations whose graphs are at bounded distances from planes of "slopes" $\omega \in \mathbb{R}^d$:

Theorem 47.1 (de la Llave et. al.) *For every $\omega \in \mathbb{R}^d$, there exists a solution of (47.1) such that*

$$\sup_{k\in\mathbb{Z}^d} |x_k - \omega \cdot k| < \infty.$$

The method of proof is very similar to the proof of the Aubry-Mather Theorem presented in GCchapter. One considers the analog of CO sequences, called Birkhoff configurations by these authors. In complete analogy to the CO sequences, they satisfy:

$$x_{k+j} + l \geq x_k,\ \forall k \in \mathbb{Z}^d \quad \text{or} \quad x_{k+j} + l \leq x_k,\ \forall k \in \mathbb{Z}^d$$

The analog to the set of CO sequences of rotation number ω, which we denoted by CO_ω in GCchapter is:

$$\mathcal{B}_\omega = \{\boldsymbol{x} \mid \ \boldsymbol{x} \text{ is Birkhoff and } \sup_{k\in\mathbb{Z}^d} |x_k - k\cdot\omega| < \infty\}$$

In a way analogous to the proof of Theorem 15.1, one shows that the gradient flow of W (that these authors, justifiably, call the heat flow) preserves order among configurations and is suitably periodic, so that the set $\mathcal{B}_\omega$ is invariant under the flow. The same argument as in the proof of Theorem 15.1 is then used to show that W must have a critical point inside $\mathcal{B}_\omega$. So, as in the classical Aubry-Mather Theorem, one not only finds solutions that have asymptotic slope ω, but these solutions have strong order properties, expressed here in terms of nonintersection: they are Birkhoff.

B*. PDE's

As Equation (47.1) suggests, the above theory smells of discretized PDE's. It is therefore not too surprising that the same kind of methods can be applied to certain PDE problems. The main ingredients necessary are some translation invariance and a heat flow that satisfies a comparison principle $u > v \Rightarrow \phi^t u > \phi^t v$, which occurs in parabolic PDE's. The method can be applied (see de la Llave (1999)) to the following PDE situations, to obtain solutions whose graphs are at bounded distance from planes with prescribed slopes, and have nonintersection properties:

$$\Delta u + V'(x,u) = 0 \tag{47.2}$$

where $V(x+e, u+\ell) = V(x,u)\ \forall x \in \mathbb{R}^d, u \in \mathbb{R}, e \in \mathbb{Z}^d, \ell \in \mathbb{Z}$.

$$\sum_{i=1}^{k} L_i^2 + V'(x,u) = 0 \tag{47.3}$$

where L_i are $\mathbb{Z}^d$ periodic vector fields satisfying Hörmander's hypoellipticity conditions and V is as in the previous case.

$$(-\Delta)^{1/2} u + V'(x,u) = 0 \tag{47.4}$$

with V as above. de la Llave (1999) also looks at the following PDE:

$$\begin{aligned} \Box\, u = u_{tt} - u_{xx} &= -V(u) + f(x,t) \\ u(x+1,t) = u(x,t+T) &= u(x,t) \end{aligned} \tag{47.5}$$

where the function f also has the periodicity:

$$f(x+1,t) = f(x,t+T) = f(x,t). \tag{47.6}$$

We say that the real number T is of constant type if its continued fraction expansion is bounded. For instance, noble numbers are of constant type.

Theorem 47.2 (de la Llave) *Let T be a number of constant type, let $f \in L^2$ satisfy (47.6) and let $V : \mathbb{R} \to \mathbb{R}$ satisfy*
(i) $0 < \alpha \leq V' \leq \beta$ where α is any positive number and β only depends on T (in an explicit manner)
(ii) $|V''(x)| \leq K$
Then there exists a weak solution $u \in L^2$ to Equation (47.5) . Moreover, if $f \in H^r$ and $V \in C^{r+2}$ has small enough C^{r+2} norm, then there is a solution $u \in H^r$ of (47.5) which is unique in a ball in H^r around the origin.

The method of proof is different from that of the above PDE's, but still involves a variational approach.

48.* Monotone Recurrence Relations

Angenent (1990) proposes a generalization of twist maps of the annulus to maps of $\mathbb{S}^1 \times \mathbb{R}^N$ which are defined by solving a recurrence relation:

$$\Delta(x_{k-l}, \dots, x_{k+m}) = 0 \tag{49.1}$$

which generalizes $\partial_2 S(x_{k-1}, x_k) + \partial_1 S(x_k, x_{k+1}) = 0$ in twist maps, where $k = l = 1$. The function Δ is required to satisfy the conditions:

a) *monotonicity* $\Delta(x_{-l}, \dots, x_{+m})$ is a non decreasing function of all the x_k except possibly for $k = 0$. Moreover, it is strictly increasing in the variables x_{-l} and x_m.

b) *periodicity* $\Delta(x_{k-l}, \dots, x_{k+m}) = \Delta(x_{k-l} + 1, \dots, x_{k+m} + 1)$

c) *coerciveness* $\lim_{x_l \to \pm\infty} \Delta(x_{-l}, \dots, x_m) = \lim_{x_m \to \pm\infty} \Delta(x_{-l}, \dots, x_m) = \pm\infty$

Under these conditions, Angenent calls (49.1) a *monotone recurrence relation*. Conditions a) and c) imply that one can solve for x_{k+m} in terms of a given $(x_{k-l}, \dots, x_{k+m-1})$. Hence this defines a map $F_\Delta : (x_{k-l}, \dots, x_{k+m-1}) \mapsto (x_{k-l+1}, \dots, x_{k+m})$ from $\mathbb{R}^{l+m}$ to itself. Condition b) implies that this maps descends to a map on $\mathbb{S}^1 \times \mathbb{R}^{l+m-1}$. Hence the N above is $N = l + m - 1$.

The notion of CO configurations, rotation number and partial order on sequences etc... of Chapter AM and GCchapter are still entirely valid here, since the variables x_k are 1 dimensional (Angenent also calls CO sequences Birkhoff). An interesting notion that Angenent (1990) introduces, inspired by PDE methods, is that of sub– or supersolution of the monotone recurrence relation (49.1) : $\underline{\boldsymbol{x}}$ is a *subsolution* if $\Delta(\underline{x}_{k-l}, \dots, \underline{x}_{k+m}) \leq 0,\ \forall k \in \mathbb{Z}$ and a *supersolution* if $\Delta(\underline{x}_{k-l}, \dots, \underline{x}_{k+m}) \geq 0,\ \forall k \in \mathbb{Z}$.

Theorem 49.1 (Angenent) *Let $\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}$ be sub– and supersolutions respectively, which are ordered: $\underline{\boldsymbol{x}} \leq \overline{\boldsymbol{x}}$. Then there is at least one solution of (49.1) , say $\boldsymbol{x}$, for which $\underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}$ holds.*

Using this theorem (whose proof is simple), Angenent (1990) is able to generalize a theorem of Hall (1984), itself a generalization of the Aubry-Mather theorem: if a twist map of the annulus, which is *not necessarily* area preserving, has a (m, n)-periodic orbit, then it must have a CO (m, n)-periodic orbit. If the map is also area preserving, this implies, taking limits, the existence of CO orbits of all rotation numbers. Analogously, Angenent proves that if there is an orbit of F_Δ with rotation number $\omega \in \mathbb{R}$, then F_Δ must also have a CO orbit of rotation number ω.

Suppose that two solutions $\boldsymbol{x}$ and $\boldsymbol{w}$ of (49.1) "exchange rotation numbers" in the sense that:

$$\lim_{n \to +\infty} x_k/k \geq \omega_1 \geq \lim_{n \to -\infty} w_k/k$$

and

$$\lim_{n \to +\infty} w_k/k \leq \omega_0 \leq \lim_{n \to -\infty} x_k/k$$

holds for some $\omega_0 \leq \omega_1$. Then Angenent proves that there must be CO orbits of any rotation number $\omega \in [\omega_0, \omega_1]$. Moreover this exchange of rotation numbers condition implies chaos: the topological entropy $h_{top}(F_\Delta) > 0$, in that there is a compact invariant set semi conjugate to a Bernouilli shift. This also generalizes shadowing results of Hall (1989) and Mather (1991a). Angenent proves a few other interesting results for the map F_Δ.

49.* Anti–Integrable Limit

MacKay & Meiss (1992) explore the existence of Aubry-Mather sets (as well as many other possible configurations) close to the anti-integrable limit, where the potential of a standard like map becomes all powerful. Consider a family F_ϵ of symplectic twist maps of $T^*\mathbb{T}^n$ given by the generating functions:

$$S_\epsilon(\boldsymbol{q},\boldsymbol{Q}) = \epsilon T(\boldsymbol{q},\boldsymbol{Q}) + V(\boldsymbol{q})$$

where, for simplicity, we can assume

$$T(\boldsymbol{q},\boldsymbol{Q}) = \frac{1}{2}(\boldsymbol{Q}-\boldsymbol{q})^2,$$

although many more general T's can be considered. As usual, orbits of F_ϵ correspond to solutions of

$$\partial_2 S_\epsilon(\boldsymbol{q}_{k-1},\boldsymbol{q}_k) + \partial_1 S_\epsilon(\boldsymbol{q}_k,\boldsymbol{q}_{k+1}) = 0. \tag{49.2}$$

Even though F_0 is not defined, it is perfectly acceptable to set $\epsilon = 0$ in Formula (49.2) . This is called the *anti-integrable limit* , a notion that seems to have appeared independently in Aubry & Abramovici (1990) and Tangerman & Veerman (1991). The force of this concept is that the solutions of (49.2) at $\epsilon = 0$ are perfectly understood: they are simply allocations of $\boldsymbol{q}_k$ to one of the critical points of V: (49.2) is just $dV(\boldsymbol{q}_k) = 0$ when $\epsilon = 0$. If V is a Morse function, it has finitely many critical points modulo $\mathbb{Z}^n$ and they are all nondegenerate. This has the following consequence:

Theorem 49.2 (MacKay–Meiss) *Any solution $\boldsymbol{q}(0)$ of (49.2) for $\epsilon = 0$ continues to a solution $\boldsymbol{q}(\epsilon)$ when ϵ is small.*

Proof. Rewrite the infinite system of equations (49.2) in the form

$$G(\epsilon,\boldsymbol{q}) = 0$$

where $G : \mathbb{R}\times X \to (\mathbb{R}^n)^{\mathbb{Z}}$ is given by $G(\epsilon,\boldsymbol{q}) = \partial_2 S_\epsilon(\boldsymbol{q}_{k-1},\boldsymbol{q}_k) + \partial_1 S_\epsilon(\boldsymbol{q}_k,\boldsymbol{q}_{k+1})$ and X is the Banach space of sequences such that $\sup_k \|\boldsymbol{q}_k - \boldsymbol{q}_k(0)\| < \infty$. The Implicit Function Theorem on Banach spaces (see Lang (1983)) applies here to find, for small ϵ, a $\boldsymbol{q}(\epsilon)$ such that $G(\epsilon,\boldsymbol{q}(\epsilon)) = 0$ as long as $\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))$ is invertible. But this is indeed the case since

$$\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))_k = V''(\boldsymbol{q}_k(0))$$

so that $\frac{\partial G}{\partial \boldsymbol{q}}(0,\boldsymbol{q}(0))$ is an infinite block diagonal matrix with the $n\times n$ diagonal blocks $V''(\boldsymbol{q}_k(0))$ all invertible and uniformly bounded. Indeed these matrices are chosen among a finite set, since $\boldsymbol{q}_k(0)$ is necessarily a critical point of V, of which there are finitely many mod $\mathbb{Z}^n$, by the assumption that V is Morse. □

One can simultaneously continue compact sets of stationary solutions from the anti-integrable limit. Such sets can be quite complicated, since the set of all stationary configurations of the anti-integrable limit can be seen as a shift on as many symbols as there are critical points. In particular, one can find invariant Cantor sets for F_ϵ. One can also get orbits with all rotation vectors $\omega \in \mathbb{R}^n$. To do so, consider the anti–integrable stationary solution $\boldsymbol{q}(0)$ which is such that $\boldsymbol{q}_0(0)$ is at some arbitrarily chosen critical point of V and

$$\boldsymbol{q}_k(0) = k[\omega] + \boldsymbol{q}_0(0),$$

where $[\omega]$ is the integer part of the vector ω. Each $\boldsymbol{q}_k(0)$ is thus on the same critical point as $\boldsymbol{q}_0(0)$, but translated by the integer vector $[\omega]$. Since $|\boldsymbol{q}_k(0) - k\omega| < \sqrt{n}$, $\boldsymbol{q}(0)$ has rotation vector ω. Now use Theorem 49.2 to continue this to an orbit of F_ϵ with rotation vector ω. One can also continue simultaneously all anti–integrable solutions as the above with rotation vectors in a compact set: they themselves form a compact set.
Even though this seems almost too easy, the anti-integrable limit is a very useful concept in order to understand the spectrum of all possible dynamics of symplectic twist maps . It is fair to say that, to this date, the least understood cases are those that are neither close to integrable nor to anti–integrable.

50.* Mather's Theory of Minimal Measures

We now come to Mather's theory of existence and regularity of minimizers. This theory is quite general: it covers a wide class of convex Lagrangian systems on tangent bundles of arbitrary compact manifolds. Note that similar, but less developed theories were created by Bangert (1989) in the setting of minimal geodesics on compact manifolds and Katok (1992) in the setting of perturbations of integrable symplectic twist maps . There is no doubt that Mather's theory could be worked out for general symplectic twist maps . Even now, the correspondences between Lagrangian systems and symplectic twist maps given in Chapter 6 (see in particular the Bialy-Polterovitch suspension theorem 40.1) should allow an ample transfer of Mather's results to the symplectic twist maps case.

The lesson we get from Mather's work is that, yes, minimizers in general manifolds behave very much like those on the circle (the realm of the classical Aubry-Mather theory), in that they satisfy a graph property. The bad news is that minimizers may be much scarcer than in the circle case: Hedlund (1932) had already constructed a Riemannian metric on $\mathbb{T}^3$ (a setting encompassed by Mather's) which is very small along 3 non intersecting geodesics which generate $H_1(\mathbb{T}^3)$. All other minimizers of a certain length are then bound to spend a good portion of their time close to these geodesics. In particular, these three geodesics are the only possible recurrent minimizers. This limits the possible rotation vectors of minimizers to these three directions only. Bangert (1989) (geodesic setting) and Mather (1991b) (Lagrangian setting) show that, in a precise sense, this is the worst case scenario: there should be at least as many rotation directions represented by minimizers as there are dimensions in $H_1(M, \mathbb{R})$. And, to end on an optimistic note, Levi (1997) construct, in this worst case scenario of Hedlund's example, "shadowing" locally minimizing orbits that spend any prescribed proportion of time close to each of the minimizers. In particular, he constructs *locally minimizing* orbits of all rotation vectors.

A*. Lagrangian Minimizers

Throughout this section and next, we consider time-periodic Lagrangian systems determined by a C^2-Lagrangian function $L : TM \times \mathbb{S}^1 \to \mathbb{R}$, where M is a compact manifold given a Riemannian metric g. Remember (see Appendix 1 or SG and Chapter 6) that extremals of the action

$$A(\gamma) = \int_a^b L(\gamma, \dot\gamma, t)dt$$

satisfies the Euler-Lagrange equations $\frac{d}{dt}\frac{\partial L}{\partial \dot x} - \frac{\partial L}{\partial x} = 0$. Using local coordinates these equations yield a first order time-periodic differential equation on TM, and thus in the standard way, a vector field on $TM \times \mathbb{S}^1$. This can be viewed as the Hamiltonian vector field corresponding to the Lagrangian system, pulled back to TM by the Legendre transformation. Since $TM \times \mathbb{S}^1$ is not compact it is possible that trajectories of this vector field are not defined for all time in $\mathbb{R}$ and thus do not fit together to give a global flow ($i.e.$ an $\mathbb{R}$-action). When the flow does exist, it is called the *Euler-Lagrange (or E-L) flow* . The following quite general hypotheses are the setting of Mather (1991b) .

Mather's Hypotheses

L is a C^2 function $L : TM \times \mathbb{S}^1 \to \mathbb{R}$ that satisfies:

(a) *Convexity:* $\frac{\partial^2 L}{\partial v^2}$ is positive definite.

(b) *Completeness:* The Euler-Lagrange flow determined by L exists.

(c) *Superlinear:* $\frac{L(x,v,t)}{\|v\|} \to \infty$ *when* $\|v\| \to +\infty$.

Mather's Hypotheses are satisfied by mechanical Lagrangians, $i.e.$ those of the form

$$L(x, v, t) = \frac{1}{2} \|v\|^2 - V(x, t),$$

where the norm is taken with respect to any Riemannian metric on the manifold. (In fact, one may allow the norm to vary with time, under some conditions, see Mañẽ (1991) , page 44).

Minimizers. We know that, for twist maps, orbits on Aubry-Mather sets are minimizers in the sense of Aubry. We have also seen in INVchapter that orbits on KAM tori are minimizers for symplectic twist maps . These are natural reasons to look for minimizers in convex Lagrangian systems. Lagrangian minimizers are defined in a way analogous to the discrete case. If $\tilde{M}$ is a covering space of M (see Appendix 2 or TOPO), L lifts to a real valued function (also called L) defined on $T\tilde{M} \times \mathbb{S}^1$. A curve segment $\gamma : [a, b] \to \tilde{M}$ is called a *$\tilde{M}$-minimizing segment* or an *$\tilde{M}$-minimizer* if it minimizes the action among all absolutely continuous curves $\beta : [a, b] \to \tilde{M}$ which have the same endpoints as γ. A curve $\gamma : \mathbb{R} \to \tilde{M}$ is also called a *minimizer* if $\gamma_{|[a,b]}$ is a minimizer for all $[a, b] \subset \mathbb{R}$. When the domain of definition of a curve is not explicitly given it is assumed to be $\mathbb{R}$. In practice, the two main covering spaces that we will consider are the universal cover (in next section) and the universal abelian cover (in this section, see Appendix 2 or TOPO for the definitions of these covering spaces).

A fundamental theorem of *Tonelli* (see Mather (1991b) or Mañẽ (1991)) implies that if L satisfies Mather's Hypotheses, then given $a < b$ and two distinct points $x_a, x_b \in \tilde{M}$ there is always a minimizer γ with $\gamma(a) = x_a$ and $\gamma(b) = x_b$. Moreover such a γ is automatically C^2 and satisfies the Euler-Lagrange equations (this uses the completeness of the E-L flow). Hence its differential $d\gamma(t) = (\gamma(t), \dot{\gamma}(t))$ yields a solution $(d\gamma(t), t)$ of the E-L flow.

Minimizers vs. hyperbolicity. There is a general principle, first unveiled by Morse in Riemannian geometry, which ties the index of the second derivative of the action of a segment of geodesic to the number of conjugate points this segment has. In terms of more general Lagrangian systems, this number can be formulated as a certain rotation index (the Maslov index) of Lagrangian subspaces under the differential of the flow along an orbit segment (see Duistermaat (1976)). If the orbit is hyperbolic, the Lagrangian tangent subspace can be chosen to be the unstable manifold. A strong illustration of this occurs in the realm of symplectic twist maps where Aubry et. al. (1991) found a striking correspondence between (local) minimizers and hyperbolicity, which we now sketch. We first need some definitions.

Given a stationary configuration $\boldsymbol{q}$ for a symplectic twist map on $T^*\mathbb{T}^n$ ($i.e.$ the $\boldsymbol{q}$ coordinate of an orbit of the map), we can define the Hessian or second derivative of the action as in the proof of 14.2 :

$$\nabla^2 W(\boldsymbol{q}) = \begin{pmatrix} \ddots & \ddots & & & & & \\ \beta_0 & \alpha_1 & \beta_1 & & & & \\ & \beta_1 & \alpha_2 & \beta_2 & & & \\ & & \beta_2 & \alpha_3 & \ddots & & \\ & & & \ddots & \ddots & \beta_{q-1} & \\ & & & & \beta_{q-1} & \alpha_q & \beta_q \\ & & & & & \ddots & \ddots \end{pmatrix}$$

an infinite matrix, tridiagonal by blocks, where

$$\alpha_k = \partial_{22} S(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k) + \partial_{11} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}), \quad \beta_k = \partial_{12} S(\boldsymbol{q}_{k-1}, \boldsymbol{q}_k).$$

If $\boldsymbol{q}$ is a local minimizer, $i.e.$ minimizes W locally on any finite segment, then the spectrum of $\nabla^2 W(\boldsymbol{x})$ is positive. We say that $\boldsymbol{q}$ has a *phonon gap* if moreover Spec$(\nabla^2 W(\boldsymbol{q})) \in [a, \infty),\ a > 0$. An invariant set has a phonon gap a if each of the orbits it contains does, and if their phonon gaps are all greater or than a.

Theorem 50.1 (Aubry–Baesens–MacKay) *Let Λ be a closed invariant set for a symplectic twist map of $\mathbb{R}^{2n}$ and Λ' be the associated set of critical sequences for W. Suppose that $\partial_{12} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$, $(\partial_{12} S)^{-1}(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$, $\partial_{11} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ and $\partial_{22} S(\boldsymbol{q}_k, \boldsymbol{q}_{k+1})$ are all bounded for $\boldsymbol{q} \in \Lambda', k \in \mathbb{Z}$. Then Λ is uniformly hyperbolic if and only if Λ has a phonon gap.*

B*. Ergodic Theory

Most of the material surveyed in this subsection can be found in Hasselblat & Katok (1995) . We start by motivating this theory by the following trivial remark: if F is a map of $T^*\mathbb{T}^n$ and $\phi(\boldsymbol{z}) = \pi(F(\boldsymbol{z})) - \pi(\boldsymbol{z})$ ($\pi : T^*\mathbb{T}^n \to \mathbb{T}^n$ is the canonical projection) then, *when it exists*:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(F^k(\boldsymbol{z})) = \lim_{n\to\infty} \frac{\pi(F^n(\boldsymbol{z})) - \pi(F(\boldsymbol{z}))}{n} = \rho_F(\boldsymbol{z}),$$

the rotation vector of $\boldsymbol{z}$ under F. The expression $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(F^k(\boldsymbol{z}))$ for a general continuous function ϕ is called the *time average* of ϕ. Hence the rotation vector of a point, when it exists, is the time average of a specific ϕ. The relevance of this is the following:

Theorem 50.2 (**Birkhoff's Ergodic Theorem**) *Let $F : (X, \mu) \to (X, \mu)$ be a measure preserving transformation for a Borel measure μ on a space X, and $\phi \in L^1(X, \mu)$. Then the time average $\phi_T(\boldsymbol{z})$ of ϕ:*

$$\phi_T(\boldsymbol{z}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(F^k(\boldsymbol{z}))$$

exists for $\mu - a.e.$ $\boldsymbol{z}$*. Moreover, if* $\mu(X) < \infty$, $\int_X \phi_T d\mu = \int_X \phi d\mu$.

Remember that a Borel measure on a topological space is one whose sigma-algebra of measurable sets is generated by the open sets. That F is measure preserving means $\mu(F^{-1}(A)) = \mu(A)$ for any Borel subset of X. An immediate corollary of Birkhoff's theorem is (one needs to compactify $T^*\mathbb{T}^n$ with a point at ∞):

Corollary 50.3 *Let* F *be a volume preserving map (eg. symplectic) of* $\mathbb{T}^n$*. The rotation vector* ρ_F *is defined on a subset of full Lebesgue measure of* $T^*\mathbb{T}^n$*.*

It turns out that the Lebesgue measure is only one of the many measures that a symplectic map F preserves. Take $\boldsymbol{z} \in T^*\mathbb{T}^n$ to be a N-periodic point of F, for instance, and let :

$$\eta = \frac{1}{N}\sum_{k=1}^{N} \delta_{F^k(\boldsymbol{z})}$$

where the Dirac measure $\delta_{\boldsymbol{w}}$ is the (Borel) probability measure concentrated at the point $\boldsymbol{w}$ ($\delta_{\boldsymbol{w}}(A)$ is 1 if $\boldsymbol{w} \in A$ and it is 0 if not). Since $\delta_{F^k(\boldsymbol{z})}(F^{-1}(A)) = \delta_{F^{k+1}(\boldsymbol{z})}(A)$, η is invariant under F. One of the many differences between η and the Lebesgue measure is their supports. In general, the *support* of a Borel measure μ is defined as:

$$\text{Supp } \mu = \{\boldsymbol{z} \in X \mid \mu(U) > 0 \quad \text{whenever} \quad \boldsymbol{z} \in U,\ U \quad \text{open} \quad \}$$

Clearly, the support of the measure constructed above is the orbit of the periodic point $\boldsymbol{z}$, whereas the support of the Lebesgue measure is $T^*\mathbb{T}^n$. Hence, the support of invariant measures is another way to conceptualize invariant sets. Let $F : X \to X$ be continuous. Then the support of any F-invariant Borel measure μ is closed, F-invariant and its complement has zero μ–measure. If $\mu(X) < \infty$, Poincaré's Recurrence Theorem implies that $\text{Supp } \mu$ is contained in the set of $F-$recurrent points. In fact, $\boldsymbol{z} \in \text{Supp } \mu \Rightarrow \boldsymbol{z} \in \omega(\boldsymbol{z}) \in \text{Supp } \mu$. Hence, to find recurrent orbits in a dynamical system, as we have been doing in this book, one can look for invariant measures.

Coming back to rotation vectors, and the measure η supported on a periodic orbit, the rotation vector $\rho_F(\boldsymbol{z})$ not only exist $\eta - a.e.$, but it is constant on $\text{Supp } \eta$. In fact, it can easily be checked that the time average ϕ_T is constant on $\text{Supp } \eta$ for any function $\phi \in L^1(T^*\mathbb{T}^n, \eta)$: the measure η is ergodic.

Definition 50.4 An F-invariant probability measure μ on a space X is *ergodic* if it satisfies one of the following equivalent properties:
1) Every F-invariant set has has μ measure 0 or 1.
2) If $\phi \in L^1(X, \mu)$ is F-invariant then ϕ is constant a.e..
3) The time average ϕ_T equals the *space average.* $\int \phi d\mu$ μ–a.e.

In terms of support, if μ is ergodic then F has an orbit in $\text{Supp } \mu$ which is dense in that support. Hence ergodicity relates to topological transitivity. The Lebesgue measure may never be ergodic for twist maps: whenever we have a chain of elliptic islands, it comprises an invariant set which is not of full Lebesgue measure. On the other hand, twist maps do have plenty of ergodic measures. We have seen above the example of a measure η supported on periodic orbits. More generally, *Aubry-Mather sets can be defined as supports* *of ergodic measures*, pull-back of measures on $\mathbb{S}^1$ invariant under circle diffeomorphisms. Indeed, take the set $\pi(M)$ in Theorem AMthmpropertiesam: it is the omega limit set $\Omega(T)$ for a circle diffeomorphism T. Now, pick $x \in \Omega(T)$ and take the weak* limit of the probability measures $\mu_N = \frac{1}{2N-1}\sum_{-N}^{N} \delta_{T^k(x)}$: it defines an ergodic measure for T, and its pull back by π is ergodic for F with support the Aubry-Mather set M. [the *weak* limit* is defined by $\mu_n \xrightarrow{*} \mu$ iff $\int_X \phi\mu_n \to \int_X \phi\mu$ for all continuous ϕ].

Hence our main objects of study in this book, periodic orbits and Aubry-Mather sets, are all supports of ergodic probability measures, part of the larger set $\mathcal{M}_F$ of all F-invariant Borel probability measures.

Remark 50.5 The existence of an ergodic measure with rotation vector (as defined by the space average) ω *does* guarantee the existence of at least one orbit with that rotation vector (the support of the measure is not empty, and the time average is constant on it). This is *not* the case if the measure is not ergodic.

If X is a compact metric space, it turns out that the set $\mathcal{M}$ of all Borel probability measures is convex and compact under the weak* topology. Moreover $\mathcal{M}_F$ itself is a compact and convex subset of $\mathcal{M}$ for this topology. A theorem of convex analysis (Krei-Millman) says that $\mathcal{M}_F$ is then in the convex hull of its *extreme points* : those $\mu \in \mathcal{M}_F$ which cannot be written as $t\mu_1 + (1-t)\mu_2$ for two distinct $\mu_1, \mu_2 \in \mathcal{M}_F$. Finally, the extreme points are all *ergodic* measures. We will see in the next subsection that there is a strong correspondence between the (strict) convexity of a certain projection of $\mathcal{M}_F$ and the Aubry-Mather theorem.

As we will see in next section, Mather (1991b) , (1993) considers measures that are invariant under the Euler-Lagrange (E-L) flow instead of a symplectic twist map . In the light of the suspension theorem of Bialy-Polterovitch (Chapter ham), his setting encompasses a large class of symplectic twist maps . All the statements that we made above are valid for E-L flows on $T^*\mathbb{T}^n$ provided one compactifies $T^*\mathbb{T}^n$ (as Mather does) in order to use the compactness of the space of E-L-invariant probability measures.

C*. Minimal Measures

For a more detailed exposition the reader is urged to consult Mather (1991b) or Manẽ (1991) . There is also a very nice survey of this theory in the beginning of Mather (1993) . Given a E-L invariant probability measure with compact support μ on $TM \times \mathbb{S}^1$, one can define its *rotation vector* $\rho(\mu)$ as follows: let $\beta_1, \beta_2, \dots, \beta_n$ be a basis of $H^1(M)$ and let $\lambda_1, \dots, \lambda_n$ be closed one-forms with $[\lambda_i] = \beta_i$ in DeRham cohomology.(11) We refer the reader uncomfortable with (co)homology to Appendix 2 or TOPO and urge her/him to read through this section thinking of the case $M = \mathbb{T}^n$, taking $[\lambda_i] = [dx_i]$, as a basis for $H^1(\mathbb{T}^n) \simeq \mathbb{R}^n$, where $(x_1, \dots, x_n)$ are angular coordinates on $T^*\mathbb{T}^n$. Define the i^{th} component of the rotation vector $\rho(\mu)$ as

$$\rho_i(\mu) = \int \lambda_i d\mu.$$

Note that this integral makes sense when one looks at λ_i as inducing a function from $TM \times \mathbb{S}^1$ to $\mathbb{R}$ by first projecting $TM \times \mathbb{S}^1$ onto TM, and then treating the form as a function on TM that is linear on fibers. The rotation vector does depend on the choice of basis β_i, but because these 1–forms are closed, $\rho_i(\mu)$ does not depend on the choice of representative λ_i with $[\lambda_i] = \beta_i$. Since the rotation vector is dual to forms, it can be

(11) When homology and cohomology coefficients are unspecified they are assumed to be $\mathbb{R}$, so the notation $H_1(M)$ means $H_1(M;\mathbb{R})$, etc.

viewed as an element of $H_1(M)$. In the case $M = \mathbb{T}^n$, one can check that, if $\gamma(0)$ is a generic point of an *ergodic* measure μ, the natural definition of *rotation vector of a curve* γ coincides with $\rho(\mu)$:

$$\rho_i(\gamma) = \lim_{b-a\to\infty} \frac{\tilde{\gamma}_i(b) - \tilde{\gamma}_i(a)}{b-a} = \lim_{b-a\to\infty} \frac{1}{b-a}\int_{d\gamma|_{[a,b]}} dx_i = \int dx_i d\mu = \rho_i(\mu)$$

where $\tilde{\gamma}$ is a lift of γ to $\mathbb{R}^n$ and the second equality uses the Ergodic Theorem (again, dx_i is seen as a function $TM \times \mathbb{S}^1 \to \mathbb{R}$). This prompts the following formula for the (i^{th} coordinate of the) rotation vector of a curve $\gamma : \mathbb{R} \to M$ for a general manifold M:

$$\rho_i(\gamma) = \lim_{b-a\to\infty} \frac{1}{b-a}\int_{d\gamma|_{[a,b]}} \lambda_i,$$

if the limit exists. As before, if $\gamma(0)$ is a generic point for an ergodic measure μ, $\rho(\gamma)$ exists and coincides with $\rho(\mu)$. Next we define the *average action* of a E-L invariant probability on $TM \times \mathbb{S}^1$:

$$A(\mu) = \int L d\mu,$$

$i.e.$ the space average of L, which equals, when μ is ergodic, to its time average along μ-a.e. orbit γ:

$$A(\mu) = \lim_{b-a\to\infty} \frac{1}{b-a}\int_a^b L(\gamma, \dot{\gamma})dt.$$

The set of E–L invariant probability measures, denoted by $\mathcal{M}_L$, is a convex set in the vector space of all measures, as we have seen in the previous subsection (It is also compact for the weak* topology if, as Mather does, one compactifies TM). and the extreme points of $\mathcal{M}_L$ are the ergodic measures (see Mañe (1987)). Now consider the map $\mathcal{M}_L \to H_1(M) \times \mathbb{R}$ given by:

$$\mu \mapsto (\rho(\mu), A(\mu)).$$

This map is trivially linear and hence maps $\mathcal{M}_L$ to a convex set U_L whose extreme points are images of extreme points of $\mathcal{M}_L$, $i.e.$ images of ergodic measures. Mather shows, by taking limits of measures supported on long minimizers representing rational homology classes, that for each ω, there exists an invariant (but not necessarily ergodic) measure μ such that $\rho(\mu) = \omega$ and $A(\mu) < \infty$.[12] Since L is bounded below, the action coordinate is bounded below on U_L. Hence we can define a map $\beta : H_1(M) \to \mathbb{R}$ by

$$\beta(\omega) = \inf\{A(\mu) \mid \mu \in \mathcal{M}_L, \rho(\mu) = \omega\},$$

which is bounded below and convex: the graph of β is the boundary of U_L. We say that a probability measure $\mu \in \mathcal{M}_L$ is a *minimal measure* if the point $(\rho(\mu), A(\mu))$ is on the graph of β. Hence, an extreme point $(\omega, \beta(\omega))$ of $graph(\beta)$ corresponds to at least one minimal ergodic measure of rotation vector ω. It turns out that if μ is minimal, μ-a.e. orbit lifts to a E-L minimizer in the universal abelian cover $\overline{M}$ of M (whose deck transformation group is $H_1(M;\mathbb{Z})/torsion$, see Appendix 2 or TOPO). Conversely, if μ is an ergodic probability measure whose support consists of $\overline{M}$-minimizers, then μ is a minimal measure.

Hence, each time we prove the existence of an extreme point $(\omega, \beta(\omega))$, we find at least one recurrent orbit of rotation vector ω which is a $\overline{M}$-minimizer.

---

[12]The impatient reader may be tempted to proclaim, from this fact, the existence of orbits of all rotation vectors. Alas, as we noted in Remark 50.5, we can guarantee that the rotation vector of orbits in the support of a measure μ are equal to $\rho(\mu)$ only when μ is ergodic

Another important property of β is that it is *superlinear*, i.e $\frac{\beta(x)}{\|x\|} \to \infty$ when $\|x\| \to \infty$. We motivate this in the simple case where $L = \frac{1}{2}\|\dot{x}\|^2 - V(x)$ and $\|\cdot\|$ comes from the Euclidean metric on the torus. If μ is any invariant probability measure, then

$$
\begin{aligned}
A(\mu) &= \int L \, d\mu \geq \int \left(\frac{\|\dot{x}\|^2}{2} - V_{max} \right) d\mu \\
&\geq \frac{1}{2} \left| \int \dot{x} \, d\mu \right|^2 - V_{max} \\
&= \frac{1}{2} |\rho(\mu)|^2 - V_{max}
\end{aligned}
\tag{50.1}
$$

where we used the Cauchy-Schwarz inequality for the second inequality. So we see that in this particular, but important, case β grows at least quadratically with the rotation vector. The superlinearity of β implies the existence of many extreme points for $graph(\beta)$ (although in most cases still too few, as we will see in the next subsection). Indeed, this growth condition implies that β's graph cannot have flat, or linear domains going to infinity. Any point $(\omega, \beta(\omega))$ is part of at least one linear domain of $graph(\beta)$, which we call S_c, where the index c denotes the "slope" (normal vector) of the supporting hyperplane whose intersection with U_L is exactly the convex and flat domain S_c. [Since c acts linearly on homology classes ω to give the equation $c \cdot \omega = a$ of S_c, it can be seen as an element of first cohomology.] Let X_c be the projection on $H_1(M)$ of S_c. The sets X_c are compact and convex domains which "tile" the space $H_1(M)$. Extreme points of X_c are projections of extreme points of S_c. Hence there are infinitely many such extreme points, and infinitely many outside any compact set. Their convex hull is $H_1(M)$, and in particular, they must span $H_1(M)$ as a vector space. Since these extreme points are the rotation vectors of minimal ergodic measures, we have found that

Theorem 50.6 *There exist at least countably many minimal ergodic measures and at least* $n = \dim H_1(M)$ *of them with distinct rotation directions.*

In particular there are at least n rotation directions represented by minimal measures for a E-L flow on $T^*\mathbb{T}^n$. We will see in Hedlund's example that this lower bound *is* attained by some systems. Finally, the *generalized Mather sets* are defined as

$$M_c = Support(\mathcal{M}_c),$$

where $\mathcal{M}_c$ is the set of all minimal measures whose rotation vectors lies in X_c. Let $\pi : TM \times \mathbb{S}^1 \to M \times \mathbb{S}^1$ denote the projection. Mather's main result in Mather (1991b) is the following theorem.

Theorem 50.7 **(Mather's Lipschitz Graph Theorem)** *For all* $c \in H^1(M)$, M_c *is a compact, non-empty subset of* $TM \times \mathbb{S}^1$. *The restriction of* π *to* M_c *is injective. The inverse mapping* $\pi^{-1} : \pi(M_c) \to M_c$ *is Lipschitz.*

In the case $M = \mathbb{T}^n$, Mather proves that, when they exist, KAM tori coincide with the sets M_c (see also Katok (1992) for some related results in the symplectic twist maps context). The proof of the Lipschitz Graph Theorem (see Mather (1991b) or Manẽ (1991)), which is quite involved, uses a curve shortening argument: if curves in $\pi(M_c)$ were too close to crossing transversally, one could "cut corners" and, because of recurrence, construct a closed curve with lesser action than A_{min}. This argument by surgery is reminiscent of the proof of Aubry's Fundamental Lemma in Chapter AM.

Remark 50.8 An important special case is that of *autonomous* systems (*i.e.* with time independent L). In this case, one can discard the time component and view M_c as a compact subset of TM. In this case, Mather's theorem implies that M_c is a Lipschitz graph for the projection $\pi : TM \to M$. To see this, suppose that two curves $x(t)$ and $y(t)$ in $\pi(M_c)$ have $x(0) = y(s)$ for some s. Mather's theorem rules out immediately the possibility that s is an integer, unless $x = y$ is a periodic orbit. For a general s, consider the curve $z(t) = y(t+s)$. Then, $\dot z(t) = \dot y(t+s)$ and, by time-invariance of the Lagrangian, $(z(t), \dot z(t))$ is a solution of the E-L flow. It has same average action and rotation vector as $(y, \dot y)$ and hence it is also in M_c. But then $z(0) = x(0)$ is impossible, by Mather's theorem, unless $\dot z(0) = \dot y(s) = \dot x(0)$ and thus, by uniqueness of solutions of ODEs, $x(t) = y(t+s)$.
By using Theorem 40.1 , one can translate the results of Mather to the realm of symplectic twist maps (see Exercise 50.9) and deduce the existence of many invariant sets that are graphs over the base and are made of minimizers. As noted before, one could also redo all of Mather's theory in the setting of symplectic twist maps (see Katok (1992), who considered the near integrable case).

D*. Examples and Counterexamples

Recovering past results. When Mather's function β (see previous section) is strictly convex, each point on $graph(\beta)$ is an extreme point and there are ergodic minimal measures (and hence minimal orbits) of all rotation vectors. One can prove that this is true when $M = \mathbb{S}^1$, and Mather (1991b) shows how his Lipschitz Graph Theorem implies the classical Aubry-Mather Theorem, by taking a E-L flow that suspends the twist map. The fact that M_c is a graph nicely translates into the fact that orbits in an Aubry-Mather set are cyclically ordered: as pointed out by Hall (1984), the CO property corresponds to trivial braiding of the suspended orbit, itself guaranteed by the graph property.

The graph of β is also strictly convex when L is a Riemannian metric on $\mathbb{T}^2$, and hence there are minimal geodesics of all rotation vectors for any metric on the torus. This was known by Hedlund (1932), who had basically worked out the same results as Aubry and Mather in that setting, albeit in a different language. [See Bangert (1988) for a unified approach of the two theories.] Hence one could hope, as a generalization of the Aubry-Mather theorem, that β is strictly convex for any Lagrangian systems satisfying Mather's hypotheses. This statement is false as we will see in the following examples.

Examples of gaps in the rotation vector spectrum of minimizers for Lagrangian on $\mathbb{T}^2$. Take $L : T\mathbb{T}^2 \to \mathbb{R}$, given by $L(x, \dot x) = \|\dot x - X\|^2$ where X is a vector field on $\mathbb{T}^2$. The integral curves x of X are automatically E-L minimizers since $L \equiv 0$ on these curves. Manẽ (1991) chooses the vector field X to be a (constant) vector field of irrational slope multiplied by a carefully chosen function on the torus which is zero at exactly one point q. The integral flow of X has the rest point $q(t) = q$, and all the other solutions are dense on the torus. The flow of X (and its lift to $T\mathbb{T}^2$ by the differential) has exactly two ergodic measures: one is the Dirac measure supported on $(\boldsymbol{q}, 0)$, with zero rotation vector, the other is equivalent to the Lebesgue measure on $\mathbb{T}^2$ and has nonzero rotation vector, say ω. Mañe checks that $\beta^{-1}(0)$ (trivially always an X_c) is the interval $\{\lambda\,\omega \mid \lambda \in [0,1]\}$, and that no ergodic measure has a rotation vectors strictly inside this interval. Thus the Mather set M_0 is the union of the supports of the two above measures .

Boyland & Golé (1996a) give an example of an autonomous *mechanical* Lagrangian on $\mathbb{T}^2$ which displays a similar phenomenon, although we also show in that paper that all autonomous Lagrangian systems satisfying Mather's Hypothesis do have minimizers of all rotation *directions*. We also give in this paper a very precise description of the β function for such systems and show that the support of minimal ergodic measures have to be either a point, a suspension of a Cantor set or a torus.

Hedlund-Bangert's counterexamples. Consider in $\mathbb{R}^3$ the three nonintersecting lines given by the x-axis, the y-axis translated by $(0,0,1/2)$ and the z-axis translated by $(1/2,1/2,0)$. Construct a $\mathbb{Z}^3$- lattice of nonintersecting axes by translating each one of these by integer vectors. Take a metric in $\mathbb{R}^3$ which is the Euclidean metric everywhere except in small, nonintersecting tubes around each of the axes in the lattice. In these tubes, multiply the Euclidean metric by a positive function λ which is 1 on the boundary and attains its (arbitrarily small) minimum along the points in the center of the tubes, *i.e.* at the axes of the lattice. Because the construction is $\mathbb{Z}^3$ periodic, this metric induces a Riemannian metric on $\mathbb{T}^3$. One can show (Bangert (1989)), if λ is taken sufficiently small, that a minimal geodesic (which is a E-L minimizer in our context) can make at most three jumps between tubes. In particular, a recurrent E-L minimizer has to be one of the three disjoint periodic orbits which are the projection of the axes of the lattice. Thus there are only three rotation directions that minimizers can take in this example, or six if one counts positive and negative orientations. In terms of Mather's theory, the level sets of the function β are octahedrons with vertices $(\pm a,0,0)$, $(0,\pm a,0)$, $(0,0,\pm a)$ (we assume here that the function λ is the same around each of the tubes). Since we are in the case of a metric, one can check that β is quadratic when restricted to a line through the origin (a minimizer of rotation vector $a\omega$ is a reparameterization of a minimizer of rotation ω). Hence a set S_c is either a face, an edge or a vertex of some level set $\{\beta = b\}$, and the corresponding M_c is, respectively, the union of three, two (parameterized at same speed) or one of the minimal periodic orbits one gets by projecting the disjoint axes. Note that, instead of the function β of Mather, Bangert uses the *stable norm*. Mather's function β is a generalization of that norm.

Levi's counter–counterexample. It is important to note that the nonexistence of minimizers of a certain rotation vector ω does not mean that there are no orbits of the E-L flow that have rotation vector ω. For example, Levi (1997) has shown the existence of orbits of all rotation vectors in the Hedlund example. He construct, using some broken geodesic methods, *local* minimizers shadowing any curve made of segments (of sufficient length) of the minimizing axes and jumps between the axes. This makes for extremely rich, chaotic dynamics.

Exercise 50.9 Find hypotheses on the generating function of an symplectic twist map F which translate to Mather's hypotheses for the Lagrangian that suspends F (*Hint.* You may want to include Bialy and Polterovitch's conditions of Theorem 40.1 for F to have a convex suspension. Note that completeness of the flow is for free: F is defined everywhere.)

51.* The Case Of Hyperbolic Manifolds

We start this section with another counterexample to the strict convexity of Mather's β function. The setting is that of a metric on the two-holed torus, the simplest example of a compact hyperbolic manifold. However, we finish the section on a positive note, by quoting a result of Boyland & Golé (1996b), in which we introduce another definition of rotation vector suited to hyperbolic manifolds and show the existence of minimal orbits of *all* rotation directions for a class of Lagrangian systems on hyperbolic manifolds only slightly smaller than that considered by Mather.

A*. Hyperbolic Counterexample

Take the metric of constant negative curvature on the surface of genus 2 (the two-holed torus) which has a long neck between the two holes (see Figure 51. 1). A minimizer here is a minimizing geodesic for the hyperbolic metric. With a and b as shown, the minimal measure for the homology class $[a] + [b]$ is a linear combination of the ergodic measures supported on Γ_a and Γ_b, where Γ_a and Γ_b are the closed geodesics in the homotopy classes of a and b, respectively. Indeed, Γ_a and Γ_b must "go around" the same holes as a and b, and any closed curve that crosses the neck will be longer than the sum of the lengths of Γ_a and Γ_b. Hence $([a] + [b], \beta([a] + [b]))$ cannot be an extreme point of $graph(\beta)$.

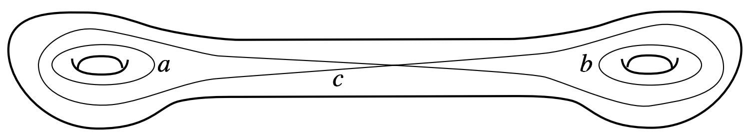


Fig. 51. 1. The surface of genus two and the loops a and b. No minimal measure with rotation vector $[a]+[b]$ can have support passing through the long neck. In particular, a curve in the homotopy class of c cannot yield a minimizer in the abelian cover.

B*. All Rotation Directions in Hyperbolic Manifolds

As the previous example shows, the notion of minimizers in the abelian cover is too restrictive, as it rules out many geodesics. Instead of working on the universal *abelian* cover, we work in the universal cover and define minimizers and rotation vectors with respect to that cover.

All manifolds of dimension n which admit a hyperbolic metric of constant curvature -1 have the Poincaré n–disk as universal covering space $\mathbb{H}^n$. Hence a hyperbolic manifold M is the quotient $\mathbb{H}^n/\pi_1(M)$ where $\pi_1(M)$ acts on $\mathbb{H}^n$ as the group of deck transformations. To visualize $\mathbb{H}^n$, assume $n = 2$, which covers any orientable surface of genus greater or equal to two. $\mathbb{H}^2$ is the usual Euclidean unit disk which is given the hyperbolic metric $\frac{dx^2+dy^2}{1-(x^2+y^2)}$. The ratio between the corresponding hyperbolic distance and the euclidean one tends exponentially to ∞ as points approach the boundary of the disk. Geodesics for the hyperbolic metric are arcs of (euclidean) circles perpendicular to the boundary $\partial\mathbb{H}^2$ of the disk.

The minimizers we consider in this section lift to curves in the *universal cover* which minimize the action between any two of their points. We also assume that the Lagrangian L satisfies Mather's hypotheses (time periodic C^2 function with (a) fiber convexity, (b) completeness of the E–L flow) except that we replace his condition (c) of superlinearity by one of superquadraticity:

(c') *superquadraticity*: There exists a $C > 0$ such that $L(x, v, t) \geq C \|v\|^2$.
(This, again, is satisfied by mechanical systems: if the potential is not positive, one can add a constant to it without changing the solutions to the system).

Theorem 51.1 (Boyland–Golé) *Let (M, g) be a closed hyperbolic manifold. Given a Lagrangian L which satisfies Hypotheses (a), (b), (c'), there are sequences k_i, κ_i, T_i in $\mathbb{R}^+$ depending only on L, with k_i increasing to infinity, such that, for any hyperbolic geodesic $\Gamma \subset \mathbb{H}^n = \tilde{M}$, there are minimizers $\gamma_i : \mathbb{R} \to \tilde{M}$ with $dist(\gamma_i, \Gamma_0) \leq \kappa_i$, $\gamma_i(\pm\infty) = \Gamma_0(\pm\infty)$, and $k_i \leq \frac{1}{d-c} dist(\gamma_i(d), \gamma_i(c)) \leq k_{i+1}$ whenever $d - c \geq T_i$.*

Theorem 51.2 (Boyland–Golé) *Let (M, g) be a closed hyperbolic manifold with geodesic flow g_t. Given a Lagrangian L which satisfies Hypotheses (a), (b), (c') with E-L flow ϕ_t, there exists sequences k_i and T_i with k_i increasing to infinity, and a family of compact, ϕ_t-invariant sets $X_i \subset \mathcal{M}$ so that for all i, (X_i, ϕ_t) is semiconjugate to $(T_1 M, g_t)$ and $k_i \leq \frac{1}{T} dist(\phi_T(\boldsymbol{x}), \phi_0(\boldsymbol{x})) \leq k_{i+1}$, whenever $T \geq T_i$ and $\boldsymbol{x} \in X_i$.*

Hence the geodesic flow and the foliation of invariant ball bundles in T^*M continues to exists, in a weak sense, in any of our general Lagrangian systems.

We now interpret Theorem 51.1 as saying that there exist minimizers of all rotation directions, with a new definition of such a concept valid only for hyperbolic manifolds. Let us first reinterpret the rotation vector on $T^*\mathbb{T}^n$ geometrically: a curve γ on $\mathbb{T}^n$ has rotation vector $\boldsymbol{v} \in \mathbb{R}^n$ if its lift $\tilde{\gamma}$ in the universal cover $\mathbb{R}^n$ is "asymptotically parallel" to the straight line supporting $\boldsymbol{v}$ and if the average of $\|\dot{\gamma}(t)\|$ over all $t \in \mathbb{R}$ is equal to $\|\boldsymbol{v}\|$ (we let the reader make these statement precise and rigorous). Now given two points on $\partial\mathbb{H}^2$, there is exactly one geodesic Γ_0 that goes to the first as $t \to -\infty$, to the other one as $t \to +\infty$. We can declare a curve γ to be asymptotically parallel to Γ_0 iff γ and Γ have same endpoints. This will insure that points of γ are always at a bounded hyperbolic distance from Γ. We also declare that the rotation vector exists iff $\tilde{\gamma}$ has the same endpoints at $\pm\infty$ as a geodesic Γ_0, and if the average $|\rho(\gamma)|$ of $\|\dot{\gamma}\|$ over $t \in \mathbb{R}$ exists, and we define the *rotation vector* to be the pair $\rho(\gamma) = (\Gamma_0, |\rho(\gamma)|)$ (average direction and average speed). In that language, Theorem 51.1 states that, given any geodesic Γ_0, there are infinitely many E-L minimizers with Γ as a rotation direction.

The naive definition of rotation vector that we just outlined has some major flaws:

1. $\rho(\gamma)$ (if it exists) does not belong to a linear space.
2. Two lifts of the same curve γ *will* have different rotation vectors.
3. Rotation direction is not constant $\mu - a.e.$ for many ergodic measures for the geodesic flow.

To remedy that, let $\pi_1(M)$, seen as deck transformation group, act on geodesics in $\mathbb{H}^2$ and declare that two geodesics are parallel iff they belong to the closure of the same $\pi_1(M)$-orbit. Consider the set of tangent vectors at all points of all the geodesics in the closure of a $\pi_1(M)$ orbit. This forms a closed subset of the unit tangent bundle of $\mathbb{H}^2$. The projection by the differential of covering map of this set on the unit tangent bundle of M is the support of a measure μ which is invariant under the geodesic flow. Because of this, Boyland (I996) defines the rotation direction of a curve to be a measure invariant under the geodesic flow, weak* limit of measures supported by geodesics joining two points of the curve. This rotation vector being defined through ergodic theory, it is constant $\mu - a.e.$ for any E-L ergodic μ. Theorem 51.2 implies the existence of minimizer of all rotation directions, in this new sense of the word.

Note that there are many more such "homotopy" directions than there are "homology" directions. For instance the "long neck" metric of Figure 51. 1 has no homology minimizer with rotation direction c, as argued in the previous subsection, but it will have infinitely many homotopy minimizers with that direction.

On the negative side, the counterexamples of Manẽ (1991) and Boyland & Golé (1996a) on $\mathbb{T}^2$ probably have counterparts on hyperbolic manifolds, even with our new definition of rotation vector and we think there is little chance to prove the existence of minimizers of all rotation vectors, even on these manifolds.

52.* Concluding Remark

So what, in the end, are the chances of finding orbits of all rotation vectors for symplectic twist maps or Lagrangian system, in say, $T^*\mathbb{T}^n$? Previous attempts at this problem yielded incomplete results. Bernstein & Katok (1987) "almost" found, for minimizing periodic orbits of symplectic twist maps close to integrable, some uniform modulus of continuity, which they hoped would unable them to take limits and get orbits with the limiting rotation vectors. In my thesis, I hoped that proving some regularity of the ghost tori (invariant set for the gradient flow of the periodic action) might enable one to do the same. This is how ghost circles came about.

One thing is clear: one cannot hope for *global* minimizers to achieve all possible rotation vectors. However, the shadowing methods to construct local minimizers of all rotation vectors of Levi (1997) on the Hedlund counterexamples indicate a possible approach to the general case. The recent work of Mather on existence of unbounded orbits (see Delshams, de la Llave & Seara (1998)and the end of INVchapter), also shows that, for general systems, hyperbolic and variational techniques can combine powerfully to construct orbits shadowing successive minimizers. One possibility to attack this problem would be to try to construct, in a manner analogous to Levi (1997), orbits shadowing the different supports of the ergodic measures which are extreme points of one generalized Mather set $\mathcal{M}_c$. Doing so, one may manage to "fill in" the corresponding set of rotation vectors X_c with rotation vectors of actual orbits, may they be local minimizers.

CHAPTER 10 or CZ

GENERATING PHASES AND SYMPLECTIC TOPOLOGY

July 15 1999

Look up Siburg's work on capacity vs symplectic twist map . What about the title of this Chapter? What about the manifold of points that only move radially? Otherwise, declare this Chapter done, after a last reading.

In Appendix 1 or SG, Section 46, we remark that the differential of a function $W : M \to T^*M$ gives rise to the Lagrangian submanifold $dW(M)$ of T^*M. As a generalization of this fact, one can construct Lagrangian submanifolds of T^*M as symplectic reductions of graphs of differentials of generating phases, which are functions on vector bundles over M.

Generating phases are the common geometric framework to the different discrete variational methods in Hamiltonian systems, including the method developped in this book. Applications of generating phases range from the search for periodic orbits to the Maslov index, symplectic capacities and singularities theory. Generating phases are a viable alternative to the use of heavy functional analytic variational methods in symplectic topology.

This chapter intends to be a basic introduction to generating phases. We first present Chaperon's method, which he used to give an alternate proof of the theorem of Conley & Zehnder (1983) . This theorem, which solved a conjecture by Arnold on the minimum number of periodic points of Hamiltonian maps of $\mathbb{T}^{2n}$, is considered by many as the starting point of symplectic topology$^{(13)}$. We then survey the abstract structure of generating phase, highlighting the common geometric frame for the symplectic twist maps method and that of Chaperon (as well as many others).

53. Chaperon's Method and the Theorem of Conley-Zehnder

Chaperon (1984) introduced a method "du type géodesiques brisées" for finding periodic orbits of Hamiltonians which did not make use of a decomposition by symplectic twist maps. This method has been the basis of later work by Laudenbach, Sikorav and Viterbo.

Until now, we have studied exact symplectic maps that come equipped with a generating function due to the twist condition. The concept of generating function is more general than this, however: we now show how an exact symplectic map of $\mathbb{R}2n$ *which is uniformly* C^1 *close to* Id may have another kind of generating

[13] In the sense that it implies that the C^0 closure of the set of symplectic diffeomorphisms is strictly included in the set of volume preserving diffeomorphisms.

function. The small time t map of a large class of Hamiltonians satisfy this condition. Hence, the time one map of these Hamiltonians can be decomposed into maps that posess this kind of generating function, leading to a new variational setting for periodic orbits. Let

$$F : \mathbb{R}2n \to \mathbb{R}2n$$
$$(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{Q}, \boldsymbol{P})$$

be an exact symplectic diffeomorphism:

$$\boldsymbol{P}d\boldsymbol{Q} - \boldsymbol{p}d\boldsymbol{q} = F^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = dS, \tag{53.1}$$

for some $S : \mathbb{R}2n \to \mathbb{R}$ (remember that all symplectic diffeomorphism of $\mathbb{R}2n$ are in fact exact symplectic. We stress *exact* symplectic here in view of our later generalization to T^*M.) The following simple lemma is crucial here.

Lemma 53.1 *Let $F : \mathbb{R}2n \to \mathbb{R}2n$ be an exact symplectic diffeomorphism. Then, if $\|F - Id\|_{C^1}$ is small enough, the map*

$$\phi : (\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{Q}, \boldsymbol{p})$$

is a diffeomorphism of $\mathbb{R}2n$.

Proof. $\boldsymbol{Q}(\boldsymbol{q}, \boldsymbol{p})$ is C^1 close to $\boldsymbol{q}$ and thus ϕ is (uniformly) C^1 close to Id, hence a diffeomorphism. □

We now show how, a way that is slightly different from the twist map case, F can be recovered from S. We define

$$\tilde{S}(\boldsymbol{Q}, \boldsymbol{p}) = \boldsymbol{p}\boldsymbol{q} + S(\boldsymbol{Q}, \boldsymbol{p}), \quad \text{where} \quad \boldsymbol{q} = \boldsymbol{q}(\boldsymbol{Q}, \boldsymbol{p});$$

then

$$d\tilde{S} = \boldsymbol{P}d\boldsymbol{Q} + \boldsymbol{q}d\boldsymbol{p} \tag{53.2}$$

and thus $\tilde{S}$ generates F, in the sense that:

$$\begin{aligned} \boldsymbol{P} &= \frac{\partial \tilde{S}}{\partial \boldsymbol{Q}}(\boldsymbol{Q}, \boldsymbol{p}) \\ \boldsymbol{q} &= \frac{\partial \tilde{S}}{\partial \boldsymbol{p}}(\boldsymbol{Q}, \boldsymbol{p}). \end{aligned} \tag{53.3}$$

Remark 53.2 Note that Id is not a symplectic twist map and thus it cannot be given a generating function in the twist map sense. One of the advantages of the present approach is that Id does have a generating function, which is

$$\tilde{S}(\boldsymbol{Q}, \boldsymbol{p}) = \boldsymbol{p}\boldsymbol{Q}$$

.

As an illustration, fixed points of F are given by the equations:

$$\boldsymbol{p} = \frac{\partial \tilde{S}}{\partial \boldsymbol{Q}} = \boldsymbol{P},$$
$$\boldsymbol{Q} = \frac{\partial \tilde{S}}{\partial \boldsymbol{p}} = \boldsymbol{q},$$

which are equivalent to the following equation:

$$d(\tilde{S} - \boldsymbol{p}\boldsymbol{Q}) = (\boldsymbol{P} - \boldsymbol{p})d\boldsymbol{Q} + (\boldsymbol{q} - \boldsymbol{Q})d\boldsymbol{p} = 0.$$

Hence have reduced the problem of finding fixed point of an exact symplectic diffeomorphism C^1 close to Id on $\mathbb{R}^{2n}$ to the one of finding critical points for a real valued function. We now apply this method to give Hamiltonian maps of $\mathbb{T}^{2n}$ a finite dimensional variational context. It can also be used for time one maps of Hamiltonians with compact support in $\mathbb{R}2n$, or Hamiltonian maps that are C^0 close to Id in a compact symplectic manifold.

Let $H : \mathbb{R}2n \times \mathbb{R}$ be a C^2 function with variables $(\boldsymbol{q}, \boldsymbol{p}, t)$. Assume H is $\mathbb{Z}^{2n}$ periodic in the variables $(\boldsymbol{q}, \boldsymbol{p})$ (i.e., H is a function on $\mathbb{T}^{2n} \times \mathbb{R}$). As in Appendix 1 or SG, we denote by $h^t_{t_0}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q}(t), \boldsymbol{p}(t))$ the solution of Hamilton's equations with initial conditions $\boldsymbol{q}(t_0) = \boldsymbol{q}, \quad \boldsymbol{p}(t_0) = \boldsymbol{p}$. By assumption, $h^t_{t_0}$ can be seen as a Hamiltonian map on $\mathbb{T}^{2n}$. We know that $h^t_{t_0}$ is exact symplectic (see Theorem 47.7). Furthermore, by compactness of $\mathbb{T}^{2n}$, *when* $|t - t_0|$ *is small,* $h^t_{t_0}$ *is* C^1 *close to* Id (the Hamiltonian vector field of a C^2 function is C^1, hence so is its flow). For $|t - t_0|$ small enough, we can apply Lemma 53.2 to get a generating function for $h^t_{t_0}$. To make this argument global, we decompose h^1 in smaller time maps (see Exercise 47.4):

$$h^1 = h^1_{\frac{N-1}{N}} \circ h^{\frac{N-1}{N}}_{\frac{N-2}{N}} \circ \ldots \circ h^{\frac{1}{N}}_{\frac{2}{N}} \circ h^{\frac{1}{N}}_0 \tag{53.4}$$

and thus, for a large enough N, h^1 can be decomposed into N maps that satisfy Lemma 53.2. [The farther h^1 is from Id, the bigger N must be.] We can then apply the following proposition to h^1:

Proposition 53.3 *Let* $F = F_N \circ \ldots \circ F_1$ *where each* F_k *is exact symplectic in* $T^*\mathbb{R}^n$, C^1 *close to* Id, *and has generating function* $\tilde{S}_k(\boldsymbol{Q}, \boldsymbol{p})$. *The fixed points of* F *are in one to one correspondence with the critical points of :*

$$\tilde{W}(\boldsymbol{Q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{Q}_N, \boldsymbol{p}_N) = \sum_{k=1}^{N} \tilde{S}_k(\boldsymbol{Q}_k, \boldsymbol{p}_k) - \boldsymbol{p}_k \boldsymbol{Q}_{k-1}$$

where we set $\boldsymbol{Q}_0 = \boldsymbol{Q}_N$.

Proof. We will use the notation

$$(\boldsymbol{P}_k, \boldsymbol{Q}_k) = F_k(\boldsymbol{q}_k, \boldsymbol{p}_k)$$

where we know from (53.3) that $\boldsymbol{P}_k$ and $\boldsymbol{q}_k$ are functions of $\boldsymbol{Q}_k, \boldsymbol{p}_k$. Then, using Equation (53.2) ,

$$\begin{aligned} d\tilde{W}(\overline{\boldsymbol{Q}}, \overline{\boldsymbol{p}}) &= \sum_{k=1}^{N} \boldsymbol{P}_k d\boldsymbol{Q}_k + \boldsymbol{q}_k d\boldsymbol{p}_k - \boldsymbol{p}_k d\boldsymbol{Q}_{k-1} - \boldsymbol{Q}_{k-1} d\boldsymbol{p}_k \\ &= \sum_{k=1}^{N-1} (\boldsymbol{P}_k - \boldsymbol{p}_{k+1}) d\boldsymbol{Q}_k + \sum_{k=2}^{N} (\boldsymbol{q}_k - \boldsymbol{Q}_{k-1}) d\boldsymbol{p}_k \\ &\quad + (\boldsymbol{P}_N - \boldsymbol{p}_1) d\boldsymbol{Q}_N + (\boldsymbol{q}_1 - \boldsymbol{Q}_N) d\boldsymbol{p}_1. \end{aligned} \tag{53.5}$$

This formula proves that $(\overline{\boldsymbol{Q}}, \overline{\boldsymbol{p}})$ is critical exactly when:

$$F_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = (\boldsymbol{q}_{k+1}, \boldsymbol{p}_{k+1}), \forall k \in \{1, \ldots, N-1\},$$
$$F_N(\boldsymbol{q}_N, \boldsymbol{p}_N) = (\boldsymbol{q}_1, \boldsymbol{p}_1),$$

that is, exactly when $(\boldsymbol{q}_1, \boldsymbol{p}_1)$ is a fixed point for F. □

As with the function W in Chapter 6, $\tilde{W}$ has the interpretation of the action of a "broken" solution of the Hamiltonian equation. This time, the jumps are both vertical *and* horizontal:

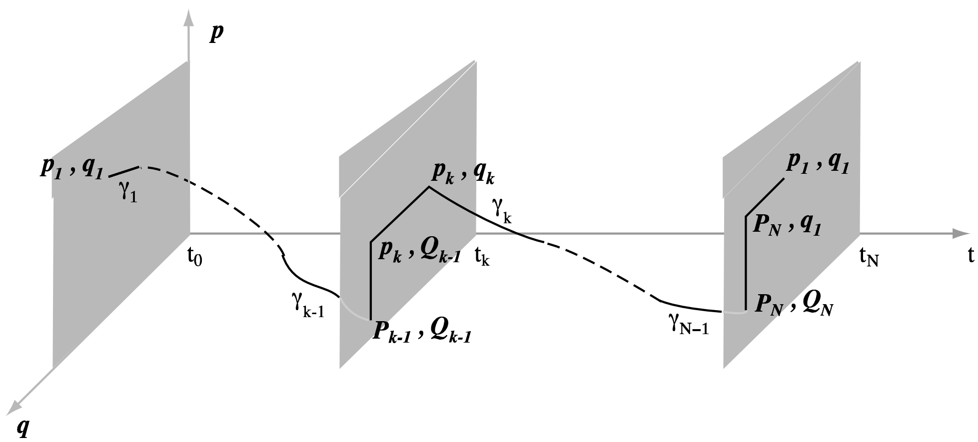


Fig. 53. 1. Interpretation of $\tilde{W}$ as the action of a "broken" solution Γ, concatenation of the solution segments γ_k and "corners" in the $t = t_k$ planes.

Each curve γ_k in Figure 53. 1 is the unique solution of Hamilton's equations starting at $(\boldsymbol{q}_k, \boldsymbol{p}_k, t_k)$ where $t_k = \frac{k-1}{N}$ and flowing for time $1/N$. Since $\tilde{S}_k(\boldsymbol{Q}_k, \boldsymbol{p}_k) = S_k(\boldsymbol{q}_k, \boldsymbol{p}_k) + \boldsymbol{p}_k\boldsymbol{q}_k$ and $S_k(\boldsymbol{q}_k, \boldsymbol{p}_k) = \int_{\gamma_k} \boldsymbol{p}d\boldsymbol{q} - Hdt$ (see Theorem 47.7), $\tilde{W}$ measures the action of the broken solution Γ:

$$\begin{aligned}\tilde{W}(\boldsymbol{Q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{Q}_N, \boldsymbol{p}_N) &= \sum_{k=1}^{N} \boldsymbol{p}_k(\boldsymbol{q}_k - \boldsymbol{Q}_{k-1}) + \sum_{k=1}^{N} \int_{\gamma_k} \boldsymbol{p}d\boldsymbol{q} - Hdt \\ &= \int_{\Gamma} \boldsymbol{p}d\boldsymbol{q} - Hdt.\end{aligned} \tag{53.6}$$

This is the definition given by Chaperon (1984) and (1989).

The following theorem, solved a famous conjecture by Arnold in the case of the Torus. It was hailed as the start of symplectic topology, as it shows that symplectic diffeomorphisms have dynamics necessarily different from that of general diffoemorphisms, or even volume preserving diffeomorphisms. The original proof of Conley and Zehnder also reduces the analysis to finite dimensions, but by truncating Fourier series of periodic orbits. Chaperon's proof avoids the functional analysis altogether.

Theorem 53.4 (Conley-Zehnder) *Let h^1 be a Hamiltonian map of $\mathbb{T}^{2n}$. Then h^1 has at least $2n+1$ distinct fixed points and at least 2^n of them if they all are nondegenerate.*

Proof. Let $\tilde{W}$ be defined as in Proposition 53.3 for the decomposition of h^1 into symplectic maps close to Id given by (53.5) . We will show that $\tilde{W}$ is equivalent to a g.p.q.i. on $\mathbb{T}^{2n}$, and hence it has the prescribed number of critical points, corresponding to fixed points of h^1. We refer the reader to Section TOPOsecgpqi for the definition and properties of generating phases that are relevant here. We first note that $\tilde{W}$ induces a function on $(\mathbb{R}2n)^N/\mathbb{Z}^{2n}$ where $\mathbb{Z}^{2n}$ acts on $(\mathbb{R}2n)^N$ by:

$$(\boldsymbol{m}_q,\boldsymbol{m}_p).(\boldsymbol{Q}_1,\boldsymbol{p}_1,\ldots,\boldsymbol{Q}_N,\boldsymbol{p}_N)=(\boldsymbol{Q}_1+\boldsymbol{m}_q,\boldsymbol{p}+\boldsymbol{m}_p,\ldots,\boldsymbol{Q}_N+\boldsymbol{m}_q,\boldsymbol{p}_N+\boldsymbol{m}_p)$$

The fact that $\tilde{W}$ is invariant under this action is most easily seen from (53.6) . Indeed, since the Hamiltonian flow is a lift from one on $\mathbb{T}^{2n}$, the curve $\gamma_k+(\boldsymbol{m}_q,\boldsymbol{m}_p,0)$ is the solution between $(\boldsymbol{q}_k+\boldsymbol{m}_q,\boldsymbol{p}_k+\boldsymbol{m}_p)$ and $(\boldsymbol{Q}_k+\boldsymbol{m}_q,\boldsymbol{P}_k+\boldsymbol{m}_p)$ starting at time $\frac{k-1}{N}$ of that flow. But

$$\int_{\gamma_k+(\boldsymbol{m}_q,\boldsymbol{m}_p,0)}\boldsymbol{p}d\boldsymbol{q}+Hdt=\int_{\gamma_k}(\boldsymbol{p}+\boldsymbol{m}_p)d\boldsymbol{q}-Hdt=\boldsymbol{m}_p(\boldsymbol{Q}_k-\boldsymbol{q}_k)+\int_{\gamma_k}\boldsymbol{p}d\boldsymbol{q}-Hdt$$

Hence the action of γ_k changes by $\boldsymbol{m}_p(\boldsymbol{Q}_k-\boldsymbol{q}_k)$ under this transformation. On the otherhand, under the same transformation, the sum $\sum_{k=1}^N\boldsymbol{p}_k(\boldsymbol{q}_k-\boldsymbol{Q}_{k-1})$ of Formula (53.6) changes by $\sum_{k=1}^N\boldsymbol{m}_p(\boldsymbol{q}_k-\boldsymbol{Q}_{k-1})$. Summing up the actions of the γ_k, these changes cancel out, from which we deduce that $\tilde{W}$ is invariant under the $\mathbb{Z}^{2n}$ action.

We now show that $\tilde{W}$ is equivalent to a g.p.q.i. over $\mathbb{T}^{2n}$. Let $E=(\mathbb{R}2n)^N\to\mathbb{R}2n$ be the bundle given by the projection map onto $(\boldsymbol{Q}_N,\boldsymbol{p}_N)$ and let $\chi:E\to E$ be the bundle diffeomorphism given by:

$$\chi(\boldsymbol{Q}_1,\boldsymbol{p}_1,\ldots,\boldsymbol{Q}_N,\boldsymbol{p}_N)=(\boldsymbol{a}_1,\boldsymbol{b}_1,\ldots,\boldsymbol{a}_{N-1},\boldsymbol{b}_{N-1},\boldsymbol{Q}_N,\boldsymbol{p}_N)$$

where

$$\begin{aligned}\boldsymbol{a}_k&=\boldsymbol{Q}_k-\boldsymbol{Q}_{k-1}\qquad(\boldsymbol{Q}_0=\boldsymbol{Q}_N)\\ \boldsymbol{b}_k&=\boldsymbol{p}_k-\boldsymbol{p}_N.\end{aligned}$$

In these new coordinates, the $\mathbb{Z}^{2n}$ action only affects $(\boldsymbol{Q}_N,\boldsymbol{p}_N)$, so that $\tilde{W}\circ\chi^{-1}$ induces a function W on $(\mathbb{R}2n)^{N-1}\times\mathbb{T}^{2n}$. We now need to show that W is in fact a g.p.q.i. Define $\tilde{W}_0$ (resp. W_0) to be the functions $\tilde{W}$ (resp. W) obtained when setting the Hamiltonian to zero. Since $\tilde{S}_k(\boldsymbol{Q}_k,\boldsymbol{p}_k)=\boldsymbol{p}_k\boldsymbol{Q}_k$ in this case, $\tilde{W}_0(\overline{\boldsymbol{Q}},\overline{\boldsymbol{p}})=\sum_{k=1}^N\boldsymbol{p}_k(\boldsymbol{Q}_k-\boldsymbol{Q}_{k-1})$ and hence a simple computation yields

$$W_0(\overline{\boldsymbol{a}},\overline{\boldsymbol{b}},\boldsymbol{Q}_N,\boldsymbol{p}_N)=\sum_{k=1}^{N-1}\boldsymbol{a}_k.\boldsymbol{b}_k$$

which, as easily checked, is quadratic nondegenerate in the fiber.

Finally, we need to check that $\frac{\partial}{\partial\boldsymbol{v}}(W_0-W)$ is bounded, where $\boldsymbol{v}=(\boldsymbol{a},\boldsymbol{b})$. It is sufficient for this to check that $d(\tilde{W}-\tilde{W}_0)$ is bounded. Using (53.5) , we obtain:

$$\begin{aligned}d(\tilde{W}-\tilde{W}_0)=&\sum_{k=1}^N(\boldsymbol{P}_k-\boldsymbol{p}_{k+1})d\boldsymbol{Q}_k+\sum_{k=1}^N(\boldsymbol{q}_k-\boldsymbol{Q}_{k-1})d\boldsymbol{p}_k\\ &-\sum_{k=1}^N(\boldsymbol{Q}_k-\boldsymbol{Q}_{k-1})d\boldsymbol{p}_k-\sum_{k=1}^N(\boldsymbol{p}_k-\boldsymbol{p}_{k+1})d\boldsymbol{Q}_k\\ =&\sum_{k=1}^N(\boldsymbol{q}_k-\boldsymbol{Q}_k)d\boldsymbol{p}_k+\sum_{k=1}^N(\boldsymbol{P}_k-\boldsymbol{p}_k)d\boldsymbol{Q}_k\end{aligned}$$

where we have set throughout $\boldsymbol{Q}_0=\boldsymbol{Q}_N,\boldsymbol{p}_{N+1}=\boldsymbol{p}_1$. Since by definition $(\boldsymbol{Q}_k,\boldsymbol{P}_k)=F_k(\boldsymbol{q}_k,\boldsymbol{p}_k)$ where $F_k=h_{\frac{k-1}{N}}^{\frac{k}{N}}$ lifts a diffeomorphism of $\mathbb{T}^{2n}$, the coefficients of the above differential must be bounded. We can conclude by applying Proposition 52.8. In fact, Proposition TOPOproptrivialgpqi is enough here. □

Remark 53.5 Since the lift of the orbits we find are closed, the orbits in $\mathbb{T}^{2n}$ are contractible. In general, one cannot hope to find periodic orbits of different homotopy classes, as the example $H_0 \equiv 0$ shows. It would be interesting, however, to study the special properties of the set of rotation vectors that orbits of h^1 may have, i.e., to find out if being Hamiltonian implies more properties on this set than those known for general diffeomorphisms of $\mathbb{T}^{2n}$.

54. Generating Phases and Symplectic Geometry

In Section TOPOsecgpqi, we define generating phases as functions $W : E \to \mathbb{R}$, where E is a vector bundle over the manifold M. We then give conditions under which lower estimates on the number of critical points of W can be obtained from the topology of M. In this section, we show how such functions give rise to Lagrangian submanifolds of T^*M, hence the adjective "generating". In particular, we show that the action function obtained either in the symplectic twist map setting or in the Chaperon approach generate a Lagrangian manifold canonically symplectomorphic to the graph of of the map F under consideration. More generally, this construction unifies the different finite, and even infinite, variational approaches in Hamiltonian dynamics.

B. Generating Phases and Lagrangian Manifolds

Let W be a differentiable function $M \to \mathbb{R}$. We have seen in Section 46.C that:

$$dW(M) = \{(\boldsymbol{q}, dW(\boldsymbol{q})) \mid \boldsymbol{q} \in M\}$$

is a Lagrangian submanifold of T^*M. Note that this manifold is a graph over the zero section 0^*_M of T^*M. Heuristicly, we would like to make it possible to similarly "generate" Lagrangian submanifolds that are not graphs. One way to do this is to add auxilliary variables and see our Lagrangian manifold as an appropriate projection in T^*M of a manifold in some bundle over M. This is what is behind the following construction.

Let $\pi : E \to M$ be a fiber bundle over the manifold M. Let $W(\boldsymbol{q}, \boldsymbol{v})$ be a real valued function on an open set $\mathcal{U} \subset E$. The derivative $\frac{\partial W}{\partial \boldsymbol{v}} : E \to E^*$ of W along the fiber of E is well defined, in the sense that if U is a chart on M and $\psi_1, \psi_2 : U \times V \to \pi^{-1}(M)$ are two local trivializations of E, and $W_1 = W \circ \psi_1,\ W_2 = W \circ \psi_2$, then

$$\Phi^* \frac{\partial W_1}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) d\boldsymbol{v} = \frac{\partial W_2}{\partial \boldsymbol{v}}(\Phi(\boldsymbol{q}, \boldsymbol{v})) d\boldsymbol{v}$$

where $\Phi = \psi_2 \circ \psi_1^{-1}$ is the change of trivialization. We assume that the map: $(\boldsymbol{q}, \boldsymbol{v}) \mapsto \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v})$ is *transverse to* 0. This means that the second derivative (in any coordinates) $(\frac{\partial^2 W}{\partial \boldsymbol{v} \partial \boldsymbol{q}}, \frac{\partial^2 W}{\partial \boldsymbol{v}^2})$ is of maximum rank at points $(\boldsymbol{q}, \boldsymbol{v})$ where $\frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0$. With this assumption, the following set of *fiber critical points* is a manifold of same dimension as M:

$$\Sigma_W = \left\{ (\boldsymbol{q}, \boldsymbol{v}) \in E \mid \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0 \right\}. \tag{54.1}$$

[For a proof of this general fact about transversality, see *eg.* the theorem p.28 in Guillemin & Pollack (1974)]
Define the map:

$$i_W : \Sigma_W \to T^*M$$
$$(\boldsymbol{q}, \boldsymbol{v}) \to \left(\boldsymbol{q}, \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v})\right)$$

Exercise CZexoimmersion shows that this is an immersion. We now show directly that this immersion is Lagrangian:

$$i_W^* \boldsymbol{p} d\boldsymbol{q} = \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}) d\boldsymbol{q} = dW\big|_{\Sigma_W}(\boldsymbol{q}, \boldsymbol{v})$$

and hence:

$$i_W^*(d\boldsymbol{q} \wedge d\boldsymbol{p}) = d^2 W\big|_{\Sigma_W} = 0.$$

We will say that W is a *generating phase* for a Lagrangian immersion $j : L \to T^*M$ if $j(L) = i_W(\Sigma_W)$.

Exercise 54.1 Show that $i_W : \Sigma_W \to T^*M$ is an immersion, *i.e.* that $Di_W\big|_{\Sigma_W}$ has full rank (*Hint.* Use the transversality condition to show that $Ker Di_W \cap T\Sigma_W = \{0\}$.)

B. Symplectic Properties of Generating Phases

We start with the trivial, but important:

Proposition 54.2 *Suppose the Lagrangian submanifold $L \subset T^*M$ is generated by a function $W : E \to \mathbb{R}$. The points in the intersection of L with the zero section 0_M^* of T^*M are in a one to one correspondance with the critical points of W.*

Proof. $i_W(\boldsymbol{q}, \boldsymbol{v})$ is in L if and only if $\frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0$. It is in 0_M^* if and only if $\frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}) = 0$. □

In TOPOsecgpqi, we find that critical points persist under elementary operations on generating phases: if $W_1 : E_1 \to \mathbb{R}$, and $W_2 : E_2 \to \mathbb{R}$ are two generating phases such that

$$W_2 \circ \Phi = W_1 + ct, \quad \text{or}$$
$$W_2(\boldsymbol{q}, \boldsymbol{v}_1, \boldsymbol{v}_2) = W_1(\boldsymbol{q}, \boldsymbol{v}_1) + f(\boldsymbol{q}, \boldsymbol{v}_2)$$

where Φ is a fiber preserving diffeomorphism and f is nondegenerate quadratic in $\boldsymbol{v}_2$, then W_1 and W_2 had the same number of critical points. The first operation is called equivalence, the second stabilization. This persistence is now geometrically explained by Proposition 54.2 and the following:

Lemma 54.3 *Two equivalent generating phases generate the same Lagrangian immersion. This is also true under stabilization.*

Proof. Let $W_2 \circ \phi = W_1 + cst$ where Φ is a fiber preserving diffeomorphism between $E_1 \to M$ and $E_2 \to M$. Writing

$$\Phi(\boldsymbol{q}, \boldsymbol{v}) = (\boldsymbol{q}, \phi(\boldsymbol{q}, \boldsymbol{v})) = (\boldsymbol{q}, \boldsymbol{v}'),$$

where $\boldsymbol{v} \to \phi(\boldsymbol{q}, \boldsymbol{v})$ is a diffeomorphism for each fixed $\boldsymbol{q}$. we have:

$$W_2(\boldsymbol{q}, \phi(\boldsymbol{q}, \boldsymbol{v})) = W_1(\boldsymbol{q}, \boldsymbol{v}) + C$$

and thus

$$\frac{\partial W_1}{\partial \boldsymbol{v}} = \left(\frac{\partial W_2}{\partial \boldsymbol{v}}' \circ \Phi\right) \cdot \frac{\partial \phi}{\partial \boldsymbol{v}}$$

This implies that $\Sigma_{W_2} = \Phi(\Sigma_{W_1})$, and we conclude the proof of the first assertion by noticing that:

$$\frac{\partial W_1}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}) = \frac{\partial W_2}{\partial \boldsymbol{q}}(\Phi(\boldsymbol{q}, \boldsymbol{v})).$$

Now let

$$W_2(\boldsymbol{q}, \boldsymbol{v}_1, \boldsymbol{v}_2) = W_1(\boldsymbol{q}, \boldsymbol{v}_1) + f(\boldsymbol{q}, \boldsymbol{v}_2)$$

where f is quadratic and nondegenerate. we have:

$$\partial W_2 / \partial \boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{v}_2 = 0 \quad \text{and} \quad \partial W_1 / \partial \boldsymbol{v}_1 = 0$$

so that $\Sigma_{W_2} = \Sigma_{W_1} \times 0_{E_2}$, where 0_{E_2} is the zero section of E_2. Moreover $\partial f / \partial \boldsymbol{q}\big|_{\{\boldsymbol{v}_2 = 0\}} = 0$ so that, at points $(\boldsymbol{q}, \boldsymbol{v}_1, 0)$ of Σ_2,

$$\left(\boldsymbol{q}, \frac{\partial W_2}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}_1, 0)\right) = \left(\boldsymbol{q}, \frac{\partial W_1}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}_1)\right).$$

□

C. The Action Function Generates the Graph of F

We examine here the twist map case, and let the reader perform the analysis for the Chaperon case in Exercise 54.4. Let M be an n–dimensional manifold and F be a symplectic twist map on $U \subset T^*M$, where U is of the form $\{(\boldsymbol{q}, \boldsymbol{p}) \in T^*M \mid \|\boldsymbol{p}\| < K\}$. Let $S(\boldsymbol{q}, \boldsymbol{Q})$ be a generating function for a lift $\tilde{F}$ of F. S can be seen as a function on some open set V of $\tilde{M} \times \tilde{M}$, diffeomorphic to $\tilde{U}$. [14] Since $\boldsymbol{P} d\boldsymbol{Q} - \boldsymbol{p} d\boldsymbol{q} = dS(\boldsymbol{q}, \boldsymbol{Q})$, we can describe the graph of $\tilde{F}$ as:

$$Graph(\tilde{F}) = \left\{\left(\boldsymbol{q}, -\frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{Q}), \boldsymbol{Q}, \frac{\partial S}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{Q})\right) \mid (\boldsymbol{q}, \boldsymbol{Q}) \in V\right\} \subset (T^*\tilde{M})^2,$$

which is canonically symplectomorphic to:

$$\left\{\left(\boldsymbol{q}, \boldsymbol{Q}, \frac{\partial S}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{Q}), \frac{\partial S}{\partial \boldsymbol{Q}}(\boldsymbol{q}, \boldsymbol{Q})\right) \mid (\boldsymbol{q}, \boldsymbol{Q}) \in V\right\} \subset T^*(\tilde{M} \times \tilde{M}).$$

One can easily check that this manifold has S as a generating phase. In other words *the generating function of a symplectic twist map F is a generating phase for the graph of $\tilde{F}$.*

We expand in more details for the more general case where $F = F_N \circ \ldots \circ F_1$ is a product of symplectic twist maps of $U \subset T^*M$. This time, the candidate for a generating phase is:

$$\tilde{W}(\overline{\boldsymbol{q}}) = \sum_{k=1}^{N} S_k(\boldsymbol{q}_k, \boldsymbol{q}_{k+1}),$$

where we do not identify $\boldsymbol{q}_{N+1}$ and $\boldsymbol{q}_1$ in any way. Then, writing

$$\boldsymbol{v} = (\boldsymbol{q}_2, \ldots, \boldsymbol{q}_N), \quad \boldsymbol{q} = (\boldsymbol{q}_1, \boldsymbol{q}_{N+1}),$$

we will show that $\tilde{W}(\boldsymbol{q}, \boldsymbol{v})$ is a generating phase for $Graph(\tilde{F}) \subset (T^*\tilde{M})^2$. Let

[14] in the case where $M = \mathbb{T}^n$, and the map is defined on all of $T^*\mathbb{T}^n$, we have $V \cong \tilde{U} \cong \mathbb{R}2n$.

$$\mathcal{U} = \left\{(\boldsymbol{q}_1, \dots, \boldsymbol{q}_k) \mid (\boldsymbol{q}_k, \boldsymbol{q}_{k+1}) \in \psi_k(\tilde{U})\right\}$$

where ψ_k is the "Legendre transformation" attached to the twist map F_k. Let $\beta : M^{N+1} \to M \times M$ be the map defined by: $(\boldsymbol{q}_1, \dots, \boldsymbol{q}_{N+1}) \to (\boldsymbol{q}_1, \boldsymbol{q}_{N+1})$ The bundle that we will consider here is:

$$\mathcal{U} \to \beta(\mathcal{U}) \subset M \times M.$$

Proposition 24.1 states that $\frac{\partial \tilde{W}}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0$ exactly when $\overline{\boldsymbol{q}} = (\boldsymbol{q}, \boldsymbol{v})$ is the $\boldsymbol{q}$ component of the orbit of $(\boldsymbol{q}_1, \boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2))$ under the successive $\tilde{F}_k$'s. This means that *the set of orbits under the successive $\tilde{F}_k$'s is in bijection with the set $\Sigma_{\tilde{W}} = \{\frac{\partial \tilde{W}}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0\}$ as defined in (54.1)* . Since this set is parametrized by the starting point of an orbit, it is diffeomorphic to U, hence a manifold.

For $\overline{\boldsymbol{q}} \in \Sigma_{\tilde{W}}$, we have:

$$\tilde{F}(\boldsymbol{q}_1, \boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2)) = \left(\boldsymbol{q}_{N+1}, \boldsymbol{P}_{N+1}(\boldsymbol{q}_N, \boldsymbol{q}_{N+1})\right)$$

but:

$$\boldsymbol{p}_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = -\partial_1 S_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = -\frac{\partial \tilde{W}}{\partial \boldsymbol{q}_1}(\boldsymbol{q}_1, \boldsymbol{q}_{N+1}, \boldsymbol{v})$$

$$\boldsymbol{P}_{N+1}(\boldsymbol{q}_N, \boldsymbol{q}_{N+1}) = \partial_2 S_N(\boldsymbol{q}_N, \boldsymbol{q}_{N+1}) = \frac{\partial \tilde{W}}{\partial \boldsymbol{q}_{N+1}}(\boldsymbol{q}_1, \boldsymbol{q}_{N+1}, \boldsymbol{v})$$

In other words, the graph of $\tilde{F}$ in $T^*\tilde{M} \times T^*\tilde{M}$ can be expressed as:

$$Graph(\tilde{F}) = \left\{ \left(\boldsymbol{q}_1, -\frac{\partial W}{\partial \boldsymbol{q}_1}(\boldsymbol{q}, \boldsymbol{v}), \boldsymbol{q}_{N+1}, \frac{\partial \tilde{W}}{\partial \boldsymbol{q}_{N+1}}(\boldsymbol{q}, \boldsymbol{v}) \right) \;\middle|\; (\boldsymbol{q}, \boldsymbol{v}) \in \Sigma_{\tilde{W}} \right\}.$$

To finish our construction, we define the following symplectic map:

$$j : \left(T^*\tilde{M} \times T^*\tilde{M}, -\Omega_{\tilde{M}} \oplus \Omega_{\tilde{M}}\right) \to \left(T^*(\tilde{M} \times \tilde{M}), \Omega_{\tilde{M} \times \tilde{M}}\right)$$
$$(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{Q}, \boldsymbol{P}) \to (\boldsymbol{q}, \boldsymbol{Q}, -\boldsymbol{p}, \boldsymbol{P}).$$

where Ω_X denotes the canonical symplectic structure on T^*X. Clearly:

$$j(Graph(\tilde{F})) = i_{\tilde{W}}(\Sigma_{\tilde{W}}),$$

that is, $\tilde{W}$ generates the Lagrangian manifold $Graph(\tilde{F})$. Note that the fixed points of F correspond to $Graph(\tilde{F}) \cap \Delta(T^*\tilde{M} \times T^*\tilde{M})$, i.e. to $\overline{\boldsymbol{q}} \in \Sigma_{\tilde{W}}$ such that $\boldsymbol{q}_1 = \boldsymbol{q}_{N+1}$ and $-\partial_1 S_1(\boldsymbol{q}_1, \boldsymbol{q}_2) = \partial_2 S_N(\boldsymbol{q}_N, \boldsymbol{q}_{N+1})$, which are critical points of $W = \tilde{W}\big|_{\{\boldsymbol{q}_1 = \boldsymbol{q}_{N+1}\}}$, as we well know.

Exercise 54.4 Show that the generating function $\tilde{W}$ of Chaperon (see Proposition 53.3) generates the graph of the Hamiltonian map $F : \mathbb{T}^{2n} \to \mathbb{T}^{2n}$. (*Hint.* If you are stuck, consult Laudenbach & Sikorav (1985)

D. Symplectic Reduction

We introduce yet another geometric point of view for the generating phase construction. We will see that if a Lagrangian manifold $L \subset T^*M$ is generated by the phase $W : E \to \mathbb{R}$, than in fact L is the symplectic reduction of the Lagrangian manifold $dW(E) \subset T^*E$. We introduce symplectic reduction in the linear case, and only sketch briefly the manifold case, refering the reader to Weinstein (1979) for more detail.

Consider V, Ω_0, be a symplectic vector space of dimension $2n$. Let C a coisotropic subspace of V. Let $\Lambda(V)$ be the set of Lagrangian subspaces of V (a Grassmanian manifold). The process of symplectic reduction gives a natural map $\Lambda(V) \to \Lambda\left(C/C^{\perp}\right)$ that we now describe. By Theorem 43.1, we know that we can find coordinates for V in which:

$$C = \{(q_1, \ldots, q_n, p_1, \ldots, p_k)\}$$

and we have $C^{\perp} = \{(q_{k+1}, \ldots, q_n)\} \subset C$. Then

$$C/C^{\perp} \simeq \{(q_1, \ldots, q_k, p_1, \ldots, p_k)\}$$

which is obviously symplectic. It is called the *reduced symplectic space* along C. We denote by Red the quotient map $C \to C/C^{\perp}$. The symplectic form Ω_C of $C/C^{\perp}$ is natural in the sense that it makes Red into a symplectic map:

$$\Omega_C(Red(\boldsymbol{v}), Red(\boldsymbol{v}')) = \Omega(\boldsymbol{v}, \boldsymbol{v}'). \tag{54.2}$$

Proposition 54.5 *Let $L \subset V$ be a Lagrangian subspace and $C \subset V$ a coisotropic subspace. Then*

$$L_C = Red(L \cap C) = L \cap C / L \cap C^{\perp}$$

is Lagrangian in $C/C^{\perp}$.

We say that L_C is *the symplectic reduction* of L along the coisotropic space C.

Proof. Formula (54.2) tells us that L_C is isotropic. We need to show that $dim L_C = \frac{1}{2} dim C/C^{\perp}$. Linear algebra tells us that:

$$dim L_C = dim(L \cap C) - dim(L \cap C^{\perp}).$$

As would be the case for any nondegenerate bilinear form, the dimensions of a subspace and that of its orthogonal add up to the dimension of the ambient space. Also, the orthogonal of an intersection is the sum of the orthogonal. Hence:

$$dim V = dim(L \cap C^{\perp}) + dim(L \cap C^{\perp})^{\perp} = dim(L \cap C^{\perp}) + dim L + dim C,$$

since $L^{\perp} = L$. Thus

$$\begin{aligned} dim L_C &= dim(L \cap C) - dim V + dim(L + C) = dim L + dim C - dim V \\ &= dim C - \frac{1}{2} dim V \end{aligned} \tag{54.3}$$

But

$$\begin{aligned} dim(C/C^{\perp}) &= dim C - dim C^{\perp} = dim C - (dim V - dim C) \\ &= 2 dim C - dim V \end{aligned} \tag{54.4}$$

We conclude that $dim L_C = \frac{1}{2} dim(C/C^{\perp})$ by putting (54.3) and (54.4) together. □

We now sketch the reduction construction in the manifold case. Let C be a coisotropic submanifold of a symplectic manifold (M, Ω). Then $TC^{\perp}$ is a subbundle of TC (that is, the fibers are of same dimension and vary smoothly) so we can form the quotient bundle $TC/TC^{\perp}$, with base C and fiber the quotient $T_{\boldsymbol{q}}C/T_{\boldsymbol{q}}C^{\perp}$ at each point $\boldsymbol{q}$ of C. It turns out that this quotient bundle can actually be seen as the tangent bundle of a certain manifold $C/C^{\perp}$, whose points are leaves of the integrable foliation $TC^{\perp}$. Moreover one can show that the naturally induced form Ω_C is indeed symplectic on $C/C^{\perp}$. Finally, we define $red : C \to C/C^{\perp}$ as the propjection. Its derivative is basically the map Red defined above. One can show that, if C intersect a Lagrangian submanifold L transversally, then $L_C = red(L)$ is an immersed symplectic manifold of $C/C^{\perp}$, which is the *reduction of L along C.*

We now apply this new point of view to the generating function construction. Let $E = M \times \mathbb{R}^N$. We show that if $L = i_W(\Sigma_W) \subset T^*M$ is generated by the generating phase $W : E \to \mathbb{R}$, then L is in fact the reduction of $dW(E) \subset T^*E$ along the coisotropic manifold $C = \{\boldsymbol{p_v} = 0\}$, where we have given T^*E the coordinate $(\boldsymbol{q}, \boldsymbol{v}, \boldsymbol{p_q}, \boldsymbol{p_v})$. This is just a matter of checking through the construction. We know that $dW(E)$ is Lagrangian in T^*E. Its intersection with C is the set:

$$\begin{aligned} dW(E) \cap C = &\left\{ (\boldsymbol{q}, \boldsymbol{v}, \boldsymbol{p_q}, \boldsymbol{p_v}) \in T^*E \;\middle|\; \boldsymbol{p_q} = \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v}),\ \boldsymbol{p_v} = \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0 \right\} \\ = &dW(\Sigma_W). \end{aligned}$$

where Σ_W is the set of fiber critical points. Since by the tranversality condition in our definition of generating phase Σ_W is a manifold, so is $dW(E) \cap C$: for any W, the map $dW : E \to T^*E$ is an embedding. The bundle $TC^{\perp}$ is the one generated by the vector fields $\frac{\partial}{\partial \boldsymbol{v}}$ and thus $C/C^{\perp}$ can be identified with $T^*M = \{(\boldsymbol{q}, \boldsymbol{p_q})\}$. The image of $dW(E) \cap C$ under the projection $red : C \to C/C^{\perp}$ is exactly $i_W(\Sigma_W) = \{(\boldsymbol{q}, \frac{\partial W}{\partial \boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{v})) \mid \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q}, \boldsymbol{v}) = 0\} = L$. Note that because $E = M \times \mathbb{R}^N$, the above argument is independent of the coordinate chosen (e.g. C is well defined.) With a little care, the argument extends to the case where E is a nontrivial bundle over M.

Exercise 54.6 Show that, in the Darboux coordinate used above, the $\boldsymbol{q}$–plane and the $\boldsymbol{p}$–plane of V both reduce to the $\boldsymbol{q}$ and $\boldsymbol{p}$–plane (resp.) of $C/C^{\perp}$.

E. Further Applications of Generating Phases

The symplectic theory of Generating Phases does not only provide a unifying packaging for the different variational approaches to Hamiltonian systems. It can also serve as the basis of symplectic topology, where invariants called *capacities* play a crucial role. Roughly speaking, capacities are to symplectic geometry what volume is to Riemannian geometry: they provide obstruction for sets to be symplectomorphic, or for sets to be squeezed inside other sets. Viterbo (1992) uses generating phases to define such capacities, in contrast to prior approaches by Gromov (1985)who uses the theory of "pseudo–holomorphic curves". The basis fo Viterbo's definition of capacity is a converse statement to Lemma 54.3:

Proposition 54.7 *If W_1 and W_2 both generate $h^t(0^*_M)$, where h^t is a Hamiltonian isotopy, then after stabilization W_1 and W_2 are equivalent.*

In view of this, Viterbo is able to define a capacity for a Lagrangian manifold L Hamiltonian isotopic to 0^*_M by choosing minimax values of a given (and hence any) generating phase for L.

In another work, Viterbo (1987) shows that a certain integer function called Maslov Index on the set of paths in the Lagrangian Grassmannian is invariant under symplectic reduction. It can be shown that the Lagrangian Grassmanian $\Lambda(V)$ has first fundamental group $\pi_1(\Lambda(V)) = \mathbb{Z}$. Very roughly, we can interpret this by saying that $\Lambda(V)$ has a "hole" and the *Maslov index* measures the number of turns a curve makes around that hole. Now let W_t be generating phases for a Hamiltonian isotopy h^t. The set $dW_T(E)$ is Lagrangian in T^*E and its reduction is graph of h^t. The Maslov Index in $\Lambda(T^*E)$ detects the change in Morse Index of the second derivative of W_t, whereas on the graph of h^t, it detects a non transverse intersection with the plane $\{(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P})\}$. This can be used to give a neat generalization to Lemma 31.2and to explain the classical relationship discovered by Morse between the index of the second variation of the action function and the number of "conjugate points"(see Milnor (1969) for the classical, Riemannian geometry case, and Duistermaat (1976) for the more general convex Lagrangian case.) Finally, we refer to Weinstein (1979) , Lecture 6, for further survey on generating phases (called Morse families there) .

Proposition CZproplagrim is 41.5, Exercise CZexoimmersion is CZexoimmersion

Appendix 1 or SG

OVERVIEW OF SYMPLECTIC GEOMETRY

(Oct 17 1999)

Action to be taken: Correct typos, decide whether to make this chapter an appendix or set it in the middle of the book (Jan 13 1999). Something weird in the header of last page!

Symplectic geometry is the language underlying the theory of Hamiltonian systems. This chapter is a short review of the main concepts, especially as they apply to Hamiltonian systems and symplectic maps in cotangent bundles. These spaces are natural when considering mechanical systems, where the base, or *configuration space* describes the position and the momentum belongs to the fiber of the cotangent bundle of the configuration space. In our optic of symplectic twist maps , one important concept studied in this chapter is that of exact symplectic map. Theorem SGhamexactsymp proves that Hamiltonian systems give rise to exact symplectic maps. We assume here some familiarity with the notions of manifold, vector bundles and differential forms. The reader who is uncomfortable with these concepts should consult any of the following references :Guillemin & Pollack (1974) or Spivak (1970) . For more on symplectic geometry and Hamiltonian systems, see Arnold (1978), Weinstein (1979), Abraham & Marsden (1985) or McDuff & Salamon (1996).

55. Symplectic Vector Spaces

In this section, we review some essentials of the linear theory of symplectic vector spaces and transformations. They will be our tools in understanding the infinitesimal behavior of symplectic maps and Hamiltonian systems in cotangent bundles. A *symplectic form* on a real vector space V is a bilinear form Ω which is skew symmetric and nondegenerate:

$$\begin{aligned} \Omega(a\boldsymbol{v} + b\boldsymbol{v}', \boldsymbol{w}) &= a\Omega(\boldsymbol{u}, \boldsymbol{w}) + b\Omega(\boldsymbol{u}', \boldsymbol{w}), \;\; (\boldsymbol{u}, \boldsymbol{u}', \boldsymbol{w} \in V, \; a, b \in \mathbb{R}). \\ \Omega(\boldsymbol{u}, \boldsymbol{w}) &= -\Omega(\boldsymbol{w}, \boldsymbol{u}) \\ \boldsymbol{u} \neq 0 &\Rightarrow \exists \boldsymbol{w} \text{ such that } \; \Omega(\boldsymbol{u}, \boldsymbol{w}) \neq 0 \end{aligned}$$

A *symplectic vector space* is a vector space V together with a symplectic form.

Example 55.1 The determinant in $\mathbb{R}^2$ is a symplectic form. More generally, the *canonical symplectic form* on $\mathbb{R}2n$, is given by:

$$\Omega_0(\boldsymbol{u}, \boldsymbol{w}) = \langle J\boldsymbol{u}, \boldsymbol{w} \rangle, \quad J = \begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix}$$

where the brackets $\langle\ ,\ \rangle$ denote the usual dot product. We will see that all symplectic vector spaces "look" like this, in particular, their dimension is always even. Usually, one denotes:

$$\Omega_0 = d\boldsymbol{p} \wedge d\boldsymbol{q} = \sum_{k=1}^{n} dp_k \wedge dq_k$$

where it is understood that dq_k, dp_k are elements of the dual basis for the coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ of $\mathbb{R}2n$. The symplectic space $(\mathbb{R}2n, \Omega_0)$ can also be interpreted as $\mathbb{R}^n \oplus (\mathbb{R}^n)^*$, equipped with the canonical symplectic form:

$$\Omega_0(\boldsymbol{a} \oplus \boldsymbol{b}, \boldsymbol{c} \oplus \boldsymbol{d}) = \boldsymbol{d}(\boldsymbol{a}) - \boldsymbol{b}(\boldsymbol{c}).$$

It is often convenient to view a bilinear form as a matrix. To do this, fix a basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V, and set:

$$A^{\Omega}_{ij} = \Omega(\boldsymbol{e}_i, \boldsymbol{e}_j)$$

Equivalently, if $\langle\ ,\ \rangle$ is the dot product associated with the basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$, then A^{Ω} is the matrix satisfying:

$$\Omega(\boldsymbol{u}, \boldsymbol{w}) = \langle A^{\Omega}\boldsymbol{u}, \boldsymbol{w}\rangle.$$

We now show that all symplectic vector spaces are isomorphic to the canonical $(\mathbb{R}2n, \Omega_0)$.

Theorem 55.2 (Linear Darboux) *If (V, Ω) is a symplectic space, one can find a basis for V in which the matrix A^{Ω} of Ω is given by* $A^{\Omega} = J = \begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix}$.

Hence, the isomorphism that sends each vector in V to its coordinate vector in the basis given by the theorem will be an isomorphism between (V, Ω) and $(\mathbb{R}^{2n}, \Omega_0)$. In classical notation, the coordinates in the Darboux coordinates are denoted by[(15)]

$$(\boldsymbol{q}, \boldsymbol{p}) = (q_1, \ldots, q_n, p_1, \ldots, p_n).$$

Proof. Since Ω is non degenerate, given any $\boldsymbol{v} \neq 0 \in V$, we can find a vector $\boldsymbol{w} \in W$ such that $\Omega(\boldsymbol{v}, \boldsymbol{w}) = -1$. In particular, the plane P spanned by $\boldsymbol{v}$ and $\boldsymbol{w}$ is a symplectic plane and the bilinear form induced by Ω on P with this basis has matrix:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{55.1}$$

Since Ω is nondegenerate on P, we must have $P^{\perp} \cap P = \{0\}$. Furthermore $V = P + P^{\perp}$, since if $\boldsymbol{u} \in V$,

$$\boldsymbol{u} - \Omega(\boldsymbol{u}, \boldsymbol{v})\boldsymbol{w} + \Omega(\boldsymbol{u}, \boldsymbol{w})\boldsymbol{v} \in P^{\perp}.$$

Ω must be nondegenerate on the $dimV - 2$ dimensional subspace $P^{\perp}$, so we can proceed by induction, and decompose $P^{\perp}$ into Ω–orthogonal planes on which the matrix of Ω is as in (55.1) . A permutation of the vectors of the basis we have found gives $A^{\Omega} = J$. □

[15]In the litterature, one also sees frequently $(\boldsymbol{p}, \boldsymbol{q})$, with $-J = \begin{pmatrix} 0 & Id \\ -Id & 0 \end{pmatrix}$ as canonical matrix.

With any bilinear form Ω on a vector space comes a notion of *orthogonal subspace* $W^\perp$ to a given subspace (or vector) W :

$$W^\perp = \{\boldsymbol{u} \in V \mid \Omega(\boldsymbol{u}, \boldsymbol{w}) = 0, \forall \boldsymbol{w} \in W\}$$

In the case of symplectic forms, the analogy with the usual notion of orthogonality can be quite misleading, as a subspace and its orthogonal will often intersect.

Exercise 55.3 Show that the linear transformation whose matrix is J in the cannonical basis is orthogonal (i.e belongs to $O(2n)$), that it satisfies $J^2 = -Id$ (i.e. J is a complex structure) as well as

$$\Omega_0(Jv, Jw) = \Omega_0(v, w)$$

(that is, J is *symplectic*, see section 57.)

Exercise 55.4 Show that a one dimensional vector subspace in a symplectic vector space is included in its own orthogonal subspace.

Exercise 55.5 Show that in a Darboux basis for a symplectic plane,

$$\Omega(\boldsymbol{v}, \boldsymbol{w}) = det(\boldsymbol{v}, \boldsymbol{w}).$$

If (q_1, p_1) are the corresponding coordinates for the plane in this basis, this determinant form is denoted by $q_1 \wedge p_1$. Show that, in Darboux coordinates for a symplectic space of dimension $2n$,

$$\Omega = \boldsymbol{q} \wedge \boldsymbol{p} = \sum_1^n q_k \wedge p_k$$

Exercise 55.6 Prove that a general skew symmetric form Ω has "normal form":

$$A^\Omega = \begin{pmatrix} 0_k & -Id_k & \\ Id_k & 0_k & \\ & & 0_l \end{pmatrix}$$

where k, l do not depend on the basis chosen.

56. Subspaces of a Symplectic Vector Space

Let V be a symplectic vector space of dimension $2n$, $W \subset V$ a subspace, and Ω_W the symplectic form restricted to W. The previous exercise shows that we can find a basis for W in which :

$$A^{\Omega_W} = \begin{pmatrix} 0_k & -Id_k & \\ Id_k & 0_k & \\ & & 0_l \end{pmatrix} \qquad dimW = 2k + l$$

In other words, (W, Ω_W) is determined up to isomorphism by k and its dimension. We will say that W is:

• *null* or *isotropic* if $k = 0$ (and $l = dimW$),

• *coisotropic* if $k + l = n$.

• *Lagrangian* if $k = 0$ and $l = n$. (i.e. W is isotropic and coisotropic.)

• *symplectic* if $l = 0$ and $k \neq 0$.

The *rank* of W is the integer $2k$.

The next theorem tells us that the qualitatively different subspaces of a symplectic space can be represented by coordinate subspaces in some Darboux coordinates.

Theorem 56.1 *A subspace W of rank $2k$ and dimension $2k+l$ in a symplectic space can be represented, in appropriate Darboux coordinates, by the coordinate plane:*

$$(q_1, \dots, q_{k+l}, p_1, \dots, p_k).$$

In particular, in some well chosen bases, an isotropic space is made entirely of q's and a coisotropic one must have at least n q's (the role of p's and q's can be reversed, of course) and a symplectic space has the same number of q's and p's.

Proof. From the definition of the rank of W, there is a subspace U of W of dimension $2k$ which is symplectic, on which we can put Darboux coordinates. $U^\perp \cap W$, the null space of Ω_W, is in the subspace $U^\perp$, which is symplectic (see Exercise 56.0.) The next lemma shows that we can complete any basis of $U^\perp \cap W$ into a symplectic basis of $U^\perp$. The union of this basis and the one in U is a symplectic basis with coordinates $(\boldsymbol{q}, \boldsymbol{p})$, in which W can be expressed as advertised. □

Lemma 56.2 *Let U be a null space in a symplectic space V Then one can complete any basis of U into a symplectic basis of V.*

Proof. Without loss of generality, V is $\mathbb{R}2n$ with its standard dot product and canonical symplectic form. Choose an orthornormal basis $(u_1, \dots, u_l)$ for U. Using the results of Exercise 55.3, the reader can easily check that JU is orthogonal to U (in the sense of the dot product) and that $(u_1, \dots, u_l, Ju_1, \dots, Ju_l)$ is a symplectic basis for $E = U \oplus JU$. From Exercise 56.4, $E^\perp \oplus E = V$ and $E^\perp$ is symplectic. We can complete the symplectic basis of E by any symplectic basis of $E^\perp$ and get a symplectic basis for V. □

As a simple consequence of Theorem 56.1, we also get:

Corollary 56.3 *If U is an isotropic subspace of a symplectic space V, one can find a coisotropic W such that $V = U \oplus W$. One can also find a Lagrangian subspace in which U is included.*

This applies in particular to Lagrangian subspaces: given any lagrangian subspace L, we can find another one L' such that $V = L \oplus L'$. In the normal coordinates of the theorem, L would be the $\boldsymbol{q}$ coordinate space, L' the $\boldsymbol{p}$ coordinate space.

Exercise 56.4 Let W be a subspace of a symplectic space V. Show that: W is symplectic $\iff W \oplus W^\perp = V \iff W^\perp$ is symplectic (Hint: see the proof of the Linear Darboux theorem).

Exercise 56.5 Show that:
W isotropic $\iff W \subset W^\perp$.
W coisotropic $\iff W^\perp \subset W$.
W is Lagrangian $\iff W$ is a maximal isotropic subspace, or minimal coisotropic subspace (for the inclusion).

Exercise 56.6 This exercise shows how symmetric matrices can be used to locally parametrize Lagrangian planes. Suppose you are given a basis $\boldsymbol{u}_1, \dots \boldsymbol{u}_n$ for a Lagrangian subspace L of $\mathbb{R}2n$. In the canonical coordinates $(\boldsymbol{q}, \boldsymbol{p})$, write $\boldsymbol{u}_k = (\boldsymbol{u}_k, \boldsymbol{w}_k)$. Let V and W be the $n \times n$ matrices whose columns are the v_k's and w_k's respectively. Suppose that L is a graph over the $\boldsymbol{q}$–plane.

(a) Show that V is invertible and that the column vectors of $\begin{pmatrix} I \\ WV^{-1} \end{pmatrix}$ form a basis for L.
(b) Show that the matrix WV^{-1} is symmetric.
(c) Deduce from this that the (Grassmanian) space of Lagrangian subspaces of $\mathbb{R}2n$ has dimension $n(n+1)/2$.

57. Symplectic Linear Maps

The Linear Darboux Theorem tells us that, up to changes of coordinates, all symplectic vector spaces are identical to $(\mathbb{R}^{2n}, \Omega_0)$. Therefore, as we define and study the transformations that preserve the symplectic form on a vector space, we need only consider the case $(\mathbb{R}^{2n}, \Omega_0)$.

Definition 57.1 A *symplectic linear map* Φ of $(\mathbb{R}^{2n}, \Omega_0)$ is a 1 to 1 linear map which leaves invariant the symplectic form:

$$\Phi^* \Omega_0 = \Omega_0, \quad \text{where } \Phi^* \Omega_0(\boldsymbol{v}, \boldsymbol{w}) := \Omega_0(\Phi \boldsymbol{v}, \Phi \boldsymbol{w}).$$

The group formed by symplectic linear maps is called the *symplectic group* and is denoted by $Sp(2n;\mathbb{R})$, or in short $Sp(2n)$. Because of the Linear Darboux Theorem, this group is naturally identified with the group of $2n \times 2n$ real matrices Φ that satisfy:

$$\Phi^t J \Phi = J, \quad J = \begin{pmatrix} 0 & -Id \\ Id & 0 \end{pmatrix} \tag{57.1}$$

Examples 57.2

(a) The group Sp(2) is exactly the group of 2×2 matrices of determinant 1.
(b) The transformation $F(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{q} + \boldsymbol{p}, \boldsymbol{p})$, with matrix $\begin{pmatrix} Id & Id \\ 0 & Id \end{pmatrix}$ is symplectic in $\mathbb{R}^{2n}$, and so is any with matrix $\begin{pmatrix} Id & A \\ 0 & Id \end{pmatrix}$, where $A^t = A$. These maps are called *completely integrable* as they preserve the n dimensional foliation of (affine) lagrangian planes $\{\boldsymbol{p} = constant\}$.
(c) A primordial example will be given by the differential of the time 1 map of Hamiltonian flows. (see Section 60.C

Symplectic linear maps have striking spectral properties:

Theorem 57.3 *Symplectic linear maps have determinant 1. If λ is an eigenvalue of a symplectic linear map, so is λ^{-1}, and they appear with the same multiplicity. If λ is a complex eigenvalue, then so are $\lambda^{-1}, \overline{\lambda}, \overline{\lambda}^{-1}$, all with the same multiplicity.*

The origin is a *hyperbolic fixed point* for a linear symplectic map when all the eigenvalues are real and distinct from ± 1. In this case the *stable and unstable manifold* (the n–dimensional union of eigen–subspaces with eigenvalues larger (resp. smaller) than 1 in absolute value) are each n dimensional. These manifolds are also Lagrangian (Exercise 57.6).

Proof. Let Φ be a symplectic map. It is not hard to see that :

$$dq_1 \wedge \ldots \wedge dq_n \wedge dp_1 \wedge \ldots \wedge dp_n = \frac{(-1)^{[n/2]}}{n!} \Omega_0 \wedge \ldots \wedge \Omega_0$$

where $[n/2]$ is the integer part of $n/2$. Since Φ preserves the right hand side of this equation, it must preserves the left hand side, i.e., the volume. Hence $\det\ \Phi = 1$. The rest of the theorem is a consequence of the fact that the characteristic polynomial $C(\lambda)$ of a symplectic transformation Φ has real coefficients and that Φ^t is similar to Φ^{-1}:

$$\Phi^t = J\Phi^{-1}J^{-1}.$$

□

Exercise 57.4 (a) Show that if a $2n \times 2n$ matrix Φ is given by its $n \times n$ block representation:

$$\Phi = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{pmatrix},$$

then Φ is symplectic if and only if $\boldsymbol{a}\boldsymbol{b}^t = \boldsymbol{b}\boldsymbol{a}^t, \quad \boldsymbol{c}\boldsymbol{d}^t = \boldsymbol{d}\boldsymbol{c}^t, \quad \boldsymbol{a}\boldsymbol{d}^t - \boldsymbol{b}\boldsymbol{c}^t = Id_n$.
(b) Show that

$$\Phi^{-1} = \begin{pmatrix} \boldsymbol{d}^t & -\boldsymbol{b}^t \\ -\boldsymbol{c}^t & \boldsymbol{a}^t \end{pmatrix}.$$

In particular, if Φ is symplectic, so are Φ^{-1} and Φ^t (this can also be shown directly from (57.1) .)

Exercise 57.5 The groups of $2n \times 2n$ real matrices $Gl(n, \mathbb{C})$ and $O(2n)$ are defined by:

$$\Phi \in Gl(n, \mathbb{C}) \Leftrightarrow \Phi J = J\Phi; \quad \Phi \in O(2n) \Leftrightarrow \Phi^t\Phi = Id$$

Show that if Φ is in any two of the groups $Sp(2n), O(2n), Gl(n, \mathbb{C})$, it is in the third. Show that, in this case, we can write:

$$\Phi = \begin{pmatrix} \boldsymbol{a} & -\boldsymbol{b} \\ \boldsymbol{b} & \boldsymbol{a} \end{pmatrix} \quad \text{with} \quad \begin{cases} \boldsymbol{a}^t\boldsymbol{b} = \boldsymbol{b}^t\boldsymbol{a} \\ \boldsymbol{a}^t\boldsymbol{a} + \boldsymbol{b}^t\boldsymbol{b} = Id \end{cases}$$

that is, the complex matrix $\boldsymbol{a} + i\boldsymbol{b}$ is in the unitary group $U(n)$.

Exercise 57.6 (a) Show that, when ± 1 is an eigenvalue of $\Phi \in Sp(n)$, it must appear with even multiplicity.
(b) Show that if λ, λ' are eigenvalues of Φ with eigenvectors $\boldsymbol{v}, \boldsymbol{v}'$ and $\lambda\lambda' \neq 1$ then $\Omega_0(\boldsymbol{v}, \boldsymbol{v}') = 0$.
(c) Deduce from (b) that, if Φ is hyperbolic, its (un)stable manifold is Lagrangian.

Exercise 57.7 Any nonsingular, real matrix Φ has the *polar decomposition*: $\Phi = PO$ where $P = (\Phi\Phi^t)^{1/2}$ is symmetric positive definite, and $O = \Phi P^{-1}$ is orthogonal. (Check this.)
(a) Show that if Φ is symplectic, then P and O are also symplectic.(*Hint.* Prove it for P by decomposing $\mathbb{R}2n$ into eigenspaces for $\Phi\Phi^t$ and using the previous exercise. Notice, in particular, that $O \in U(n)$, by Exercise 57.5.
(b) Show more generally that $(\Phi\Phi^t)^\alpha$ is symplectic for all real α, and deduce from this that $U(n)$ is a deformation retract of $Sp(2n)$.

58. Symplectic Manifolds

Let N be a differentiable manifold. A *symplectic structure* on N is a family of symplectic forms on the tangent spaces of N which depends smoothly on the base point and has a certain nondegeneracy condition. Technically, a symplectic structure is given by a closed nondegenerate differential 2–form Ω:

$$d\Omega = 0 \text{ and, for all } \boldsymbol{v} \neq 0 \in T_zM, \exists \boldsymbol{w} \in T_zM \text{ such that } \Omega(\boldsymbol{v}, \boldsymbol{w}) \neq 0.$$

Ω is called a *symplectic form* and (M, Ω) a *symplectic manifold*. A *symplectic map* or *symplectomorphism* between two symplectic manifolds (N_1, Ω_1) and (N_2, Ω_2) is a differentiable map $F : N_1 \to N_2$ such that:

$$F^* \Omega_2 = \Omega_1.$$

In other words, the tangent space at each point of a symplectic manifold is a symplectic vector space, and the differential of a symplectic map at a point is a symplectic linear map between symplectic vector spaces.

Example 58.1
(a) Once again , the canonical example is given by $(\mathbb{R}^{2n}, \Omega_0)$, where $\mathbb{R}^{2n}$ is thought of as a manifold. The tangent space at a point is identified with $\mathbb{R}^{2n}$ itself, and the form Ω_0 is a constant differential form on this manifold.
(b) Any surface with its volume form is a symplectic manifold. Symplectic maps in dimension 2 are just area preserving maps.
(c) Kähler manifolds (see McDuff & Salamon (1996)) are symplectic.
(d) Cotangent bundles are non compact symplectic manifolds (see Section 59) and time 1 maps of Hamiltonian vector fields on them are symplectic maps.

The fundamental theorem by Darboux (of which we have proven the linear version) says that locally, all symplectic manifolds are isomorphic to $(\mathbb{R}^{2n}, \Omega_0)$. See Arnold (1978), Weinstein (1979) or McDuff & Salamon (1996) for a proof of this.

Theorem 58.2 (Darboux) *Let (N, Ω) be a symplectic manifold. Around each point of N, one can find a coordinate chart $(\boldsymbol{q}, \boldsymbol{p})$ such that :*

$$\Omega = \sum_1^n dq_k \wedge dp_k := d\boldsymbol{q} \wedge d\boldsymbol{p}.$$

Hence all $2n$–dimensional symplectic manifolds are locally symplectomorphic. This is in sharp contrast with Riemannian geometry, where for example the curvature, is an obstruction for two manifolds to be locally isometric.

Submanifolds of a symplectic manifold can inherit the qualitative features of their tangent spaces: A submanifold $Z \subset (N, \Omega)$ is *(co)isotropic* if each of its tangent spaces is (co)isotropic in the symplectic tangent space of N. Hence a *Lagrangian submanifold* is an isotropic submanifold of dimension $n = \frac{1}{2} dim N$. Any curve on a surface is a Lagrangian submanifold. The 0-section and the fiber of the cotangent bundle of a manifold (see next Section) is a Lagrangian submanifold, and so is the graph of any closed differential form.

Exercise 58.3 Show the following:
(a) Any symplectic manifold has even dimension.
(b) If (N, Ω) is a $2n$ dimensional symplectic manifold, then Ω^n is a volume form .
(c) A symplectomorphism is a volume preserving diffeomorphism.

Exercise 58.4 Let (N, Ω) be a symplectic manifold and $F : N \to N$ a symplectomorphism. Show that the set *graph* F is a Lagrangian submanifold of $(N \times N, \Omega \oplus (-\Omega))$

59. Cotangent Bundles

A. Some definitions

Let M be a differentiable manifold of dimension n. Its *cotangent bundle* $T^*M \xrightarrow{\pi} M$ is the fiber bundle whose fiber T_q^*M at a point q of M is the dual to the fiber T_qM of the tangent bundle. The elements of T_q^*M are *cotangent vectors* or linear 1–forms, based at q. Given local coordinates $(q_1, \ldots, q_n)$ in a chart of M , one usually denotes a tangent vector $\boldsymbol{v}$ by:

$$\boldsymbol{v} = \sum_1^n v_k \frac{\partial}{\partial q_k}$$

where $\frac{\partial}{\partial q_k}$ denotes the tangent vector to the k th coordinate line at the point q considered. A cotangent vector p at the point q takes the form:

$$\boldsymbol{p} = \sum_1^n p_k dq_k$$

Where dq_k denotes the 1–form dual to $\frac{\partial}{\partial q_k}$:

$$dq_j(\frac{\partial}{\partial q_k}) = \delta_{jk}.$$

Once the system of coordinates $\boldsymbol{q} = (q_1, \ldots, q_n)$ is chosen, the coordinates $\boldsymbol{p} = (p_1, \ldots, p_n)$ for T_q^*M are uniquely determined, and we call them the *conjugate coordinates*. The cotangent bundle T^*M as a smooth union of the fibers T_q^*M is a differentiable manifold of dimension $2n$, with local coordinates $(\boldsymbol{q}, \boldsymbol{p})$ as presented above. More precisely, if $\boldsymbol{q} \xrightarrow{\Psi} \boldsymbol{Q}$ is a coordinate change between two charts U and V of M, then :

$$(\Psi^*)^{-1}(\boldsymbol{q}, \boldsymbol{p}) = (\boldsymbol{Q}, \boldsymbol{P}) = (\Psi(\boldsymbol{q}), (D\Psi_{\boldsymbol{q}}^t)^{-1}\boldsymbol{p})$$

is a change of coordinates in the corresponding charts $U \times \mathbb{R}^n$ and $V \times \mathbb{R}^n$ of T^*M. This law of change of coordinates is what distinguishes tangent vectors from cotangent vectors. More generally, whenever we have a (local) diffeomorphism $F : M \to N$ between two manifolds M and N, there is (locally) an induced *pull–back map*: $F^* : T^*N \to T^*M$ which can be written $F^*(\boldsymbol{q}, \boldsymbol{p}) = (F^{-1}(\boldsymbol{q}), DF_{\boldsymbol{q}}^t(\boldsymbol{p}))$ in coordinates.

Example 59.1
(a)$\mathbb{R}^{2n} \cong \mathbb{R}^n \oplus (\mathbb{R}^n)^*$ can be seen as the cotangent bundle of the manifold $\mathbb{R}^n$: this bundle is trivial, as any bundle over a contractible manifold.
(b)The cotangent bundle of $\mathbb{T}^n$ is $\mathbb{T}^n \times \mathbb{R}^n$. That $T^*\mathbb{T}^n$ is trivial is a consequence of the fact that $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$, where $\mathbb{Z}^n$ acts as a group of translations on $\mathbb{R}^n$, whose differentials are the Id. See the following exercise.

Exercise 59.2 More generally, if $M \cong \mathbb{R}^n/\Gamma$ where Γ is a group of diffeomorphisms of $\mathbb{R}^n$ acting *properly discontinuously* (i.e. around each point q of M there is a neighborhood $U(q)$ such that $U \cap (\Gamma \backslash Id)(U) = \emptyset$), then

$$T^*M \cong \mathbb{R}^{2n}/\Gamma^*$$

where Γ^* is the set of diffeomorphisms of $\mathbb{R}^{2n}$ of the form γ^*, where $\gamma \in \Gamma$.

B. Cotangent Bundles as Symplectic Manifold

We now show that there is a natural symplectic structure on T^*M. We first construct a canonical differential 1–form called the *Liouville form* which, we will prove, has the following expression in any set of conjugate coordinates:

$$\lambda = \sum_1^n p_k dq_k = \boldsymbol{p} d\boldsymbol{q}.$$

We then obtain a symplectic form by differentiating λ:

$$\Omega = -d\lambda, \quad \Omega = d\boldsymbol{q} \wedge d\boldsymbol{p},$$

the latter holding in any conjugate coordinate system.

We first present a coordinate free construction of λ. To define a 1–form on T^*M, it suffices to determine how it acts on any given tangent vector v in a fiber $T_\alpha(T^*M)$ of the tangent plane of T^*M. Since the base point α is in T^*M, it is a linear 1–form. Let $\pi : T^*M \to M$ be the canonical projection. The derivative $\pi_* : T(T^*M) \to TM$ takes a vector $\boldsymbol{v}$ to the vector $\pi_*\boldsymbol{v}$ in $T_{\pi(\alpha)}M$. We can evaluate the 1–form α on that vector, and define:

$$\lambda(\boldsymbol{v}) = \alpha(\pi_*\boldsymbol{v})$$

See Figure 59. 5

Fig. 59. 5. The Liouville form on T^*M.

We now compute λ in local, conjugate coordinates. If $(\boldsymbol{q}, \boldsymbol{p})$ are the conjugate coordinates of T^*M, we can write:

$$\alpha = \sum \alpha_k dq_k \text{ and } \boldsymbol{v} = \boldsymbol{u_q}\frac{\partial}{\partial \boldsymbol{q}} + \boldsymbol{u_p}\frac{\partial}{\partial \boldsymbol{p}}.$$

Then $\pi_*(\boldsymbol{v}) = \boldsymbol{u_q}\frac{\partial}{\partial \boldsymbol{q}}$ and $\alpha(\pi_*\boldsymbol{u}) = \sum \alpha_k v_{q_k}$ which exactly says that $\lambda = \boldsymbol{p}d\boldsymbol{q}$. □

The fact that the symplectic form Ω is *exact* ($i.e.$ the differential of another form, here λ) on a cotangent bundle enables us to single out an important class of symplectic map: one way to say that $F : T^*M \to T^*M$ is symplectic in T^*M is to say that the form $F^*\lambda - \lambda$ is closed:

$$d(F^*\lambda - \lambda) = F^*d\lambda - d\lambda = -(F^*\Omega - \Omega) = 0$$

Definition 59.3 A map $F : T^*M \to T^*M$ is *exact symplectic* if $F^*\lambda - \lambda$ is exact:

$$F^*\lambda - \lambda = dS$$

for some real valued function S on T^*M.

We will see in Section 60 that time t maps of flows arising in classical mechanics ($i.e.$ Hamiltonian flows) are all exact symplectic, and so are most of the maps in this book. Note that in $\mathbb{R}2n$, since any closed form is exact, symplectic and exact symplectic are two equivalent properties. On the other hand, the map $(x, y) \to (x, y + a), \quad a \neq 0$, of the cylinder is a good example of a map which is symplectic but not exact symplectic.

Remark 59.4 The term *exact* diffeomorphism, or even *exact symplectic* diffeomorphism is sometimes used to denote the time 1 map of a (time dependent) Hamiltonian system. We will see in Section 60 that, on cotangent bundles, these time–1 maps are indeed exact symplectic in the sense of our definition. It can be shown that the map $(\boldsymbol{q}, \boldsymbol{p}) \mapsto (\boldsymbol{q} + A\boldsymbol{p}, \boldsymbol{p}),\ A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ is exact symplectic but not isotopic to Id (true more generally whenever A is not homotopic (cannot be deformed) to I on $\mathbb{T}^2$). Hence these maps cannot be time–1 maps of Hamiltonians. Cotangent bundles are just one example, albeit the most important one, of *exact symplectic manifolds*: symplectic manifolds whose symplectic form is exact. Many facts that are true for cotangent bundles also hold for exact symplectic manifolds.

Exercise 59.5 Show that the set of exact symplectic maps forms a group under composition. In particular, show that generating functions if $G * \lambda - \lambda = S_G$ and $F^*\lambda - \lambda = S_F$ then

$$(F \circ G)^*\lambda - \lambda = d[(S_F \circ G) + S_G]$$

Exercise 59.6 Let $F : T^*\mathbb{S}^1 \to T^*\mathbb{S}^1$. The *net flux* of F through a non contractible simple closed loop C is the difference between the area above C but below $F(C)$, and the area above $F(C)$ but below C.
(a) Show that, if F is symplectic, the net flux is independent of the choice of C.
(b) Show that if F is symplectic then: F is exact symplectic $\Leftrightarrow$ F has zero net flux . In particular, an area preserving map of the annulus that has an invariant circle is automatically exact symplectic.

Exercise 59.7 Show that a map F of T^*M is exact symplectic if and only if :

$$\int_{F\gamma} \boldsymbol{p}d\boldsymbol{q} = \int_{\gamma} \boldsymbol{p}d\boldsymbol{q}$$

for all differentiable closed curve γ.

C. Notable Lagrangian Submanifolds of Cotangent Bundles

It is not hard to see that the fibers of T^*M are Lagrangian submanifolds of T^*M: in coordinates they are given by $\{\boldsymbol{q} = \boldsymbol{q}_0\}$ and hence their tangent space is of the form $\{\boldsymbol{q} = 0\}$. Likewise, the zero section 0^*_M of T^*M is Lagrangian. Another class of example will be of importance to us in Chapter CZ. Consider a function $W : M \to \mathbb{R}$. Its differential dW can be seen as a section of T^*M, i.e. a map $M \to T^*M$ whose image $dW(M)$ can be written as $\{(\boldsymbol{q}, dW(\boldsymbol{q})) \mid \boldsymbol{q} \in M\}$. A basis for the tangent space of $dW(M)$ at a point $(\boldsymbol{q}, dW(\boldsymbol{q}))$ is given by:

$$\boldsymbol{v}_k = \frac{\partial}{\partial q_k} + \sum_{j=1}^{n} \frac{\partial^2 W(\boldsymbol{q})}{\partial q_j \partial q_k} \frac{\partial}{\partial p}_j$$

It is not hard to see that:

$$\Omega(\boldsymbol{v}_k, \boldsymbol{v}_l) = \frac{\partial^2 W(\boldsymbol{q})}{\partial q_k \partial q_l} - \frac{\partial^2 W(\boldsymbol{q})}{\partial q_l \partial q_k} = 0,$$

so that $dW(M)$ *is a Lagrangian submanifold of* M. We can generalize this argument somewhat. Any 1–form α can be seen as a map from M to T^*M, so we can ask the question: for what α is $\alpha(M)$ a Lagrangian manifold ? To answer this question, one can check (Exercise 59.8) the following formula:

(59.1) $$\alpha^*\lambda = \alpha.$$

where λ is the Liouville form (the reader has to get used to the fact that we see α either as a form or a map, at our convenience. When seen as a map, α is actually an embedding of M into T^*M.)

The manifold $\alpha(M)$ is Lagrangian exactly when:

$$0 = \alpha^*\Omega = \alpha^*(-d\lambda) = -d(\alpha^*\lambda) = -d\alpha,$$

that is, exactly when α is a closed form. In particular, if the form α is exact with $\alpha = dW$, this gives another proof that $dW(M)$ is Lagrangian. W is the simplest instance of *generating function* for the Lagrangian manifold $\alpha(M) = dW(M)$ (*generating phase* or *generating phase function* is also used). We will expend on this important notion of symplectic topology in Chapter CZ.

Exercise 59.8 Verify Formula (59.1) , using local coordinates.

60. Hamiltonian Systems

A. Lagrangian Systems versus Hamiltonian systems

A lot of mechanical problems can be put in terms of a variational problem: under the *principle of least action*, trajectories are critical points of an *action functional* of the form:

$$A(\gamma) = \int_{t_0}^{t_1} L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)dt,$$

with boundary condition $\gamma(t_0) = \boldsymbol{q}_0, \gamma(t_1) = \boldsymbol{q}_1$. The function L is twice differentiable in each variables, say (absolute continuity is enough). It is called the *Lagrangian function* of the system. As this is a somewhat heuristic discussion, we will not specify here the functional space to which γ belongs. In concrete cases (say $\gamma \in C^1([t_0, t_1]), \mathbb{R}^n)$ or $C^1([t_0, t_1], M)$, or some Sobolev space of curves...), the following can be made quite rigorous.

To compute the differential of A, one applies a small variation $\delta\gamma = (\delta\boldsymbol{q}, \delta\boldsymbol{p})$ to γ, with $\delta\gamma(t_0) = \delta\gamma(t_1) = 0$. Then:

$$\delta A(\gamma) = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)\delta\boldsymbol{q} + \frac{\partial L}{\partial \dot{\boldsymbol{q}}}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)\delta\dot{\boldsymbol{q}}\right) dt.$$

performing an integration by parts on the second term of this integral, we get:

$$\delta A(\gamma) = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) \delta\boldsymbol{q} dt$$

Since this should be true for any variation $\delta\gamma$, we must have:

(60.1) $$\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = 0,$$

which is a second order differential equation in $\boldsymbol{q}$ called the *Euler- Lagrange* equations. (The plural to "equations" just refers to the fact that the dimension is usually greater than 1.) As an example, a large number of mechanical systems have a Lagrangian function of the form:

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \frac{1}{2}\|\dot{\boldsymbol{q}}\|^2 - V_t(\boldsymbol{q}).$$

("Kinetic - potential". The time dependance of V usually refers to some forcing) where $V : \mathbb{R}^n \to \mathbb{R}, f : \mathbb{R} \to \mathbb{R}^n$. The Euler–Lagrange equations for such a system are:

$$\ddot{\boldsymbol{q}} + \nabla V_t(\boldsymbol{q}) = 0$$

To solve such an O.D.E., one usually proceeds by introducing $\boldsymbol{p} = \dot{\boldsymbol{q}}$ to get a system of first order ODE's:

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{p} \\ \dot{\boldsymbol{p}} &= -\nabla V(\boldsymbol{q}). \end{aligned}$$

As we will see presently, we have just put the Lagrangian problem into a Hamiltonian form. In general, if

$$\det \frac{\partial^2 L}{\partial \dot{\boldsymbol{q}}^2} \neq 0, \tag{60.2}$$

we can introduce

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$$

to transform the Euler–Lagrange equations (60.1) into a system of first order O.D.E.'s: because of the nondegeneracy condition (60.2) , the implicit function theorem implies that, locally, we can make a change of variables :

$$\mathcal{L} : (\boldsymbol{q}, \dot{\boldsymbol{q}}) \to (\boldsymbol{q}, \boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}) \tag{60.3}$$

This is, when $\boldsymbol{q}$ is seen as a point on a manifold M, a local diffeomorphism between $T_{\boldsymbol{q}}M$ and $T^*_{\boldsymbol{q}}M$. This change of variables is called the *Legendre transformation.*[16]

Define the *Hamiltonian function* by:

$$H(\boldsymbol{q}, \boldsymbol{p}, t) = \boldsymbol{p}\dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t),$$

Where it is understood that $\dot{\boldsymbol{q}} = \dot{\boldsymbol{q}} \circ \mathcal{L}^{-1}(\boldsymbol{q}, \boldsymbol{p})$. We can compute:

$$\begin{aligned} \frac{\partial H}{\partial \boldsymbol{q}} &= \boldsymbol{p}\frac{\partial \dot{\boldsymbol{q}}}{\partial \boldsymbol{q}} - \frac{\partial L}{\partial \boldsymbol{q}} - \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\frac{\partial \dot{\boldsymbol{q}}}{\partial \boldsymbol{q}} = -\frac{\partial L}{\partial \boldsymbol{q}}, \\ \frac{\partial H}{\partial \boldsymbol{p}} &= \dot{\boldsymbol{q}} + \boldsymbol{p}\frac{\partial \dot{\boldsymbol{q}}}{\partial \boldsymbol{p}} - \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\frac{\partial \dot{\boldsymbol{q}}}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}}. \end{aligned}$$

But the Euler-Lagrange equations imply that:

$$\dot{\boldsymbol{p}} = \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = \frac{\partial L}{\partial \boldsymbol{q}} = -\frac{\partial H}{\partial \boldsymbol{q}}.$$

Combining this with the previous formula yields *Hamilton's equations*:

$$\begin{aligned} \dot{\boldsymbol{q}} &= H_{\boldsymbol{p}} \\ \dot{\boldsymbol{p}} &= -H_{\boldsymbol{q}}. \end{aligned} \tag{60.4}$$

Remark 60.1

(a) The Legendre transformation is *involutive*: it is its own inverse, in the following sense. The map

[16] In the classical literature the term Legendre transformation refers to the complete process of changing the Lagrangian L into the Hamiltonian H as shown in this section, and H is then called the Legendre transformation of L. It is grammatically less awkward to call H the Legendre *transformed* of L.

$(\boldsymbol{q}, \dot{\boldsymbol{q}}) \to (\boldsymbol{q}, \frac{\partial L}{\partial \dot{\boldsymbol{q}}} = \boldsymbol{p})$ has inverse:

$$(\boldsymbol{q}, \boldsymbol{p}) \to (\boldsymbol{q}, \frac{\partial H}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}})$$

and L is the Legendre transformed of H in the sense that:

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \boldsymbol{p}\dot{\boldsymbol{q}} - H(\boldsymbol{q}, \boldsymbol{p}, t)$$

where $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$ is given implicitly by $\frac{\partial H}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}}$.

(b) In the new coordinates, the action functional becomes:

$$A(\gamma) = \int_\gamma \boldsymbol{p} d\boldsymbol{q} - H dt$$

where γ is seen as a curve $(\boldsymbol{q}(t), \boldsymbol{p}(t), t)$ in the space $\mathbb{R}2n \times \mathbb{R}$, or $T^*M \times \mathbb{R}$.

Hamilton's equations have a natural expression in the symplectic setting. We assume now that $\boldsymbol{q}$ is in $\mathbb{R}^n$. Using the notation $H_t(\boldsymbol{q}, \boldsymbol{p}) = H(\boldsymbol{q}, \boldsymbol{p}, t)$, we can rewrite (60.4) as

$$\dot{z} = -J\nabla H_t(z) \overset{def}{:=} X_H(z, t).$$

where $\nabla H_t = \binom{H_q}{H_p}$ is the gradient of H_t with respect to the scalar product on $\mathbb{R}^{2n}$:

$$\langle \nabla H_t, \boldsymbol{v} \rangle = dH_t(\boldsymbol{v})$$

Likewise, X_H, which we call the *Hamiltonian vector field* should be seen as the *symplectic gradient* of H_t:

$$\Omega_0(X_H, \boldsymbol{v}) = \langle -J^2 \nabla H_t, \boldsymbol{v} \rangle = \langle \nabla H_t, \boldsymbol{v} \rangle = dH_t(\boldsymbol{v}).$$

This can be written using the contraction operator on differential forms:

$$i_{X_H} \Omega = dH_t$$

Exercise 60.2
(a) Compute the Legendre transformed of $L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \frac{1}{2}\langle A\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}} \rangle - V(\boldsymbol{q})$.
(b) Show that, in general, if H is the Legendre transformed of L, then

$$L_{\dot{\boldsymbol{q}}\dot{\boldsymbol{q}}} H_{\boldsymbol{pp}} = Id.$$

B. Hamiltonian Systems on a Symplectic Manifold

Motivated by the last expression that we found for the Hamiltonian vector field in $\mathbb{R}^{2n}$, we extend the definition to symplectic manifolds:

Definition 60.3 Let (N, Ω) be a symplectic manifold and $H(z, t) = H_t(z)$ be a real valued function on $N \times \mathbb{R}$. The *Hamiltonian vector field* associated with H is the (time dependent) vector field X_H defined by:

$$\Omega(X_H, \boldsymbol{v}) = dH_t(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in TM.$$

Equivalently:

$$i_{X_H} \Omega = dH_t$$

The (time dependent) O.D.E.:

$$\dot{\boldsymbol{z}} = X_H(\boldsymbol{z}, t) \tag{60.5}$$

is called *Hamilton's equations.*

In local Darboux coordinate charts (*eg.* in conjugate coordinates chart of a cotangent bundles), these equations take the form of (60.4) . If H is time independent, then (60.5) generates a (local) flow on N.If H is time dependent, then X_H generates a (local) flow in the space $N \times \mathbb{R}$, called the *extended phase space* in mechanics. Specifically, one solves the following time independent system on $N \times \mathbb{R}$:

$$\dot{\boldsymbol{z}} = X_H(\boldsymbol{z}, s)$$
$$\dot{s} = 1$$

which generates a flow ϕ^t in $N \times \mathbb{R}$ satisfying:

$$\phi^t(\boldsymbol{z}, s) = (h_s^{t+s}(\boldsymbol{z}), s + t),$$

where h_s^t is a family of C^{k-1} diffeomorphims of N , depending C^{k-1} on s and t. This is a general procedure for time dependent vector field. The diffeomorphism h_s^t is called a *Hamiltonian map* and, for each fixed s the curve $t \rightarrow h_s^t$ is a *Hamiltonian isotopy* (an *isotopy* is a smoothly varying 1–parameter family of diffeomorphisms). Another way of describing $h_s^t(\boldsymbol{z})$ is by saying that it is the unique solution $\boldsymbol{z}(t)$ of Hamilton's equation with initial condition $\boldsymbol{z}(s) = \boldsymbol{z}$. In practice, one often fixes $s = 0$ and denotes h_0^t by h^t.

The following exercise shows the one to one correspondence between time dependent vector fields and isotopies. It also shows that, even though the time 0 of a solution flow to a time dependent vector field is the Identity, the flow does not in general form a group.

Exercise 60.4 Let X_t be a vector field (not necessarily Hamiltonian) on a manifold N. Let h_s^t be the solution flow to the O.D.E. $\dot{\boldsymbol{z}} = X_t(\boldsymbol{z}), \dot{s} = 1$. Prove that:
(i) $h_s^s = Id, \ \forall s$,
(ii)$h_s^{t'} = h_t^{t'} \circ h_s^t$, so that in particular $h_s^t = h^t \circ (h^s)^{-1}$. Compute $(h_s^t)^{-1}$.
(iii) Conversely, given any (sufficiently smooth) isotopy g^t in N, with $g^0 = Id$, show that the time dependent vector field:

$$\dot{g}^t = \frac{d}{ds}\Big|_{s=0} g^{t+s} \circ (g^t)^{-1}$$

has solution $h_s^t = g^t \circ (g^s)^{-1}$.

C. Invariants of the Hamiltonian Flow

We analyze here how different objects vary under the Hamiltonian flow. If G is a function on a differentiable manifold N, and X is a vector field, we recall that the *Lie derivative* of G along X is:

$$L_X G(z) = \frac{d}{dt}\Big|_{t=0} G(\phi^t(z)) = dG(X(z))$$

where ϕ^t is the flow solution for X .

Theorem 60.5 *Let H be a time independent Hamiltonian function on (M, Ω). Then H is constant under the Hamiltonian flow it generates:*

$$L_{X_H} H = 0.$$

Proof. $L_{X_H} H = dH(X_H) = \Omega(X_H, X_H) = 0$ □

Remark 60.6 $L_{X_H} G = \Omega(X_G, X_H) = -L_{X_G} H$ is also denoted by $\{G, H\}$ and it is called the *Poisson bracket* of H and G. Hence, the poisson bracket measures how far the function G (resp. H) is from being constant along the flow of X_H (resp.X_G). When $\{H, G\} = 0$, one says that G (resp. H) is a *first integral* of the Hamiltonian flow of H (resp. G), or that the functions H and G are *in involution*. One can show (see *eg.* Arnold (1978), Abraham & Marsden (1985)) that the set of Hamiltonian vector fields form a Lie sub– algebra of the Lie algebra of vector fields on a manifold, in the sense that:

$$X_{\{H,G\}} = [X_H, X_G].$$

In particular, the poisson bracket of two functions measures how far from commuting their Hamiltonian flows are.

One can also compute how a differential form α varies along an isotopy g_t by the Lie derivative. Let us first extend the notion of Lie derivative to differential forms. If X is any vector field, we define *the Lie derivative in the direction of X* by:

$$L_X \alpha = \frac{d}{dt}\Big|_{t=0} g_t^* \alpha.$$

where g_t is the flow generated by X. At time $t \neq 0$,

$$\frac{d}{dt} g_t^* \alpha = g_t^* L_X \alpha.$$

Hence, the isotopy g_t preserves the form α whenever this Lie derivative is zero:

$$g_t^* \alpha = \alpha, \ \forall t \Longleftrightarrow L_X \alpha = 0.$$

We have the important *homotopy formula* (see *eg.* McDuff & Salamon (1996)):

$$L_X \alpha = i_X d\alpha + d(i_X \alpha) \tag{60.6}$$

and again, at time $t \neq 0$,

$$\frac{d}{dt} g_t^* \alpha = g_t^* \left(i_X d\alpha + d(i_X \alpha) \right)$$

A *symplectic isotopy* g_t on (M, Ω) is an isotopy such that g_t is a symplectic map for all t. By the homotopy formula (and the fact that a symplectic form is closed), this can be reworded:

$$g_t \text{ is a symplectic isotopy} \Longleftrightarrow L_X \Omega = 0 \Longleftrightarrow d(i_X \Omega) = 0 \tag{60.7}$$

The following theorem characterises Hamiltonian isotopies, at least in cotangent bundles (or in any exact symplectic manifold, i.e. one whose symplectic form is exact)

Theorem 60.7 *(a) On any symplectic manifold, Hamiltonian isotopies are symplectic.*
(b) On a cotangent bundle T^*M, a Hamiltonian isotopy with Hamiltonian $H(\boldsymbol{z},t)$ is also exact symplectic:

$$h^{t^*}\lambda - \lambda = h^{t^*}\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = dS_t, \quad \textit{with } S_t = \int_\gamma \boldsymbol{p}d\boldsymbol{q} - Hd\tau$$

where γ is the curve $(h^\tau(\boldsymbol{z}),\tau)$, $\tau \in [0,t]$ solution of Hamilton's equations in the extended phase space $T^*M \times \mathbb{R}$, and $\boldsymbol{z}$ is the point at which the form is evaluated.
(c) Conversely, if an isotopy g_t is exact symplectic then it is Hamiltonian, with the Hamiltonian function given by:

$$H_t = i_{X_t}\boldsymbol{p}d\boldsymbol{q} - (g_t^{-1})^*\frac{d}{dt}(S_t).$$

where $X_t(\boldsymbol{z}) = \frac{dg_t}{dt}(g_t^{-1}(\boldsymbol{z}))$.

Proof. The first assertion (a) is an immediate consequence of (60.7) : if h^t is a Hamiltonian isotopy then $i(\dot{h}_t)\Omega = dH_t$ is exact, and therefore closed. In cotangent bundles, it is also a consequence of the second assertion. We look for $\frac{d}{dt}(S_t)$ in the statement (b):

$$\frac{d}{dt}h_t^*\lambda = h_t^*\left(i_{X_H}d\lambda + d(i_{X_H}\lambda)\right) = h_t^*d(-H_t + i_{X_H}\lambda) = dh_t^*(-H_t + i_{X_H}\lambda)$$

From this we get:

$$h_t^*\lambda - \lambda = d\int_0^t h_\tau^*(-H_\tau + i_{X_H}\lambda)d\tau \stackrel{\text{def}}{=} dS_t$$

that is, h^t is exact symplectic. We leave it to the reader to rewrite the integral as the one advertised in the theorem. This finishes the proof of (b). To prove the converse (c), let g_t be an exact symplectic isotopy:

$$g_t^*\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} = dS_t$$

for some S_t differentiable in all of $(\boldsymbol{q},\boldsymbol{p},t)$. We claim that the (time dependent) vector field:

$$X_t(\boldsymbol{z}) = \frac{dg_t}{dt}(g_t^{-1}(\boldsymbol{z}))$$

whose time t is g_t, is Hamiltonian. To see this, we compute:

$$\frac{d}{dt}(dS_t) = \frac{d}{dt}g_t^*\boldsymbol{p}d\boldsymbol{q} = g_t^*L_{X_t}\boldsymbol{p}d\boldsymbol{q} = g_t^*\left(i_{X_t}d(\boldsymbol{p}d\boldsymbol{q}) + d(i_{X_t}\boldsymbol{p}d\boldsymbol{q})\right),$$

from which we get

$$i_{X_t}d\boldsymbol{q}\wedge d\boldsymbol{p} = d\left(i_{X_t}\boldsymbol{p}d\boldsymbol{q} - (g_t^{-1})^*\frac{d}{dt}(S_t)\right) = dH_t$$

which exactly means that X_t is Hamiltonian with H_t as Hamiltonian function. □

A less formal proof of (b) in the above theorem yields extra information. We follow Chapter 9 in Arnold (1978). We first prove that the vector field $(X_H,1)$ in $T^*M \times \mathbb{R}$ generates the kernel of the form $d(\boldsymbol{p}d\boldsymbol{q} - Hdt) = d\boldsymbol{p}\wedge \boldsymbol{dq} - H_{\boldsymbol{q}}d\boldsymbol{q}\wedge dt - H_{\boldsymbol{p}}d\boldsymbol{p}\wedge dt$. The matrix of this form in the (Darboux) coordinate $(\boldsymbol{q},\boldsymbol{p},t)$ is:

$$A = \begin{pmatrix} 0 & -Id & H_{\boldsymbol{q}} \\ Id & 0 & H_{\boldsymbol{p}} \\ -H_{\boldsymbol{q}} & -H_{\boldsymbol{p}} & 0 \end{pmatrix}.$$

since the upper left $2n\times 2n$ matrix is the nonsingular matrix J, A is of rank (at least) $2n$. It is easy to see that the vector $(H_{\boldsymbol{p}}, -H_{\boldsymbol{q}}, 1)$ generates its kernel.

Now, take a closed curve γ in $T^*M \times \mathbb{R}$. The image under the Hamiltonian flow of γ forms an embedded tube in $T^*M \times \mathbb{R}$. Since the tangent space to this tube at any of its point $\boldsymbol{z}$ contains the vector $X_H(\boldsymbol{z})$, the form $d(\boldsymbol{p}d\boldsymbol{q} - Hdt)$ restricted to this tube is null. As a result, because of Stokes' theorem, if γ_1 and γ_2 in $T^*M \times \mathbb{R}$ encircle the same tube of orbits of the extended flow, we must have:

$$\int_{\gamma_1} \boldsymbol{p}d\boldsymbol{q} - Hdt = \int_{\gamma_2} \boldsymbol{p}d\boldsymbol{q} - Hdt \tag{60.8}$$

since $\gamma_1 - \gamma_2$ is the boundary of a region of the tube. The form $\boldsymbol{p}d\boldsymbol{q} - Hdt$ is called the *integral invariant of Poincaré–Cartan*. As a particular case, if γ_1 is of the form (γ, t_1) and $\gamma_2 = (h_{t_1}^{t_2}\gamma, t_2)$, the form Hdt is null on these curves and hence Equation (60.8) reads:

$$\int_{\gamma_1} \boldsymbol{p}d\boldsymbol{q} = \int_{h_{t_1}^{t_2}\gamma} \boldsymbol{p}d\boldsymbol{q} \tag{60.9}$$

This last equation implies the statement (b) in Theorem 60.7: it proves that the function

$$S_t = \int_{\boldsymbol{z}_0}^{\boldsymbol{z}} h^{t*}\boldsymbol{p}d\boldsymbol{q} - \boldsymbol{p}d\boldsymbol{q} \tag{60.10}$$

is well defined, i.e. the integral does not depend on the path chosen between z_0 and z. This proves in turn that h^t is exact symplectic. □

The next theorem, due to Jacobi, runs somewhat against the title of this subsection, in the sense that we show that symplectic diffeomorphisms conserve Hamilton's equations. This property in fact characterises symplectic transformations, which are for this reason called *canonical transformations* in the classical litterature. Even though we will not need this theorem in the sequel, we include it here since it explains why symplectic geometry came to exist.

Theorem 60.8 *Let $F : (M, \omega_M) \to (N, \omega_N)$ be a diffeomorphism. Then F is symplectic if and only if for all function $H : N \to \mathbb{R}$,*

$$F_* X_{H\circ F} = X_H. \tag{60.11}$$

In this case, F conjugates the Hamiltonian flows h^t and g^t of H and $H \circ F$ respectively:

$$g^t = F^{-1} \circ h^t \circ F.$$

This holds also when H is time dependent.

Proof. Reminding the reader that by definition $F_*X(F(z)) = DF_zX(z)$ for any vector field X, we also use the notation F^*Y to mean $(F^{-1})_*Y$. It is not hard to check that the following formula holds:

$$F^* i_X \alpha = i_{F^*X} F^* \alpha \tag{60.12}$$

for any vector field X and differential form α. Coming back to our statement, we have on one hand:

$$F^* i_{X_H} \omega_N = F^* dH = dH \circ F$$

by tracking down definitions, and on the other hand,

$$F^* i_{X_H}\omega_N = i_{F^*X_H} F^*\omega_N = i_{F^*X_H}\omega_M$$

because of (60.12) and the fact that F is symplectic. This proves (60.11) . Conversally, if (60.11) holds for any H, the same kind of computation shows that,

$$i_{X_{H\circ F}} F^*\omega_N = i_{X_{H\circ F}}\omega_M$$

and since any tangent vector at a point of M is of the form $X_{H\circ F}$ for some H, we must have $F^*\omega_N = \omega_M$, i.e. F is symplectic. The conjugacy statement, a general fact about O.D.E.'s, is left to the reader, as well as checking that everything still works with time dependent systems. □

Exercise 60.9 The Lie derivative of a function can be defined, in the obvious way, along any differentiable isotopy. What fails in Theorem 60.5 when H is time dependent?

Exercise 60.10 Show that in Darboux coordinates:

$$\{H, G\} = \frac{\partial H}{\partial \boldsymbol{q}}\frac{\partial G}{\partial \boldsymbol{p}} - \frac{\partial H}{\partial \boldsymbol{p}}\frac{\partial G}{\partial \boldsymbol{q}}.$$

Exercise 60.11 Prove that the function S_t defined in (60.10) satisfies:

$$S_t(\boldsymbol{z}) = \int_\gamma \boldsymbol{p}d\boldsymbol{q} - Hdt \quad + C(\boldsymbol{z}_0, t),$$

for some C, and γ as in Theorem 60.10. (*Hint.* Apply Stokes on the appropriate surface.)

Exercise 60.12 Prove that h_s^t is exact symplectic (i.e. even for $s \neq 0$), where $h_s^t(\boldsymbol{z})$ is, as in subsection B, the solution of Hamilton's equation such that $\boldsymbol{z}(s) = \boldsymbol{z}$.

Exercise 60.13 Let H be autonomous , or of period τ. Show that $X_H(\boldsymbol{z})$ is preserved by $Dh^\tau(\boldsymbol{z})$, *i.e.* X_H is an eigenvector of Dh^τ with eigenvalue 1.

D. Symplectic Maps as Return Maps of Hamiltonian Systems

Consider a time independent Hamiltonian on $\mathbb{R}^{2n+2}$, with its standard symplectic structure $\Omega_0 = \sum_{k=0}^n dq_k \wedge dp_k$. Assume that we have a periodic trajectory γ for the Hamiltonian flow. It must then lie in an energy level $H = H_0 = H(\gamma(0))$, since H is time independent. Take any $2n+1$ dimensional open disk $\tilde{\Sigma}$ which is transverse to γ at $\gamma(0)$, and such that $\tilde{\Sigma}$ intersects γ only at $\gamma(0)$.

Such a disk clearly always exists, if γ is not a fixed point. In fact, one can assume that, in a local Darboux chart, $\tilde{\Sigma}$ is the hyperplane with equation $q_0 = 0$: this is because in the construction of Darboux coordinates, one can start by choosing an arbitrary nonsingular differentiable function as one of the coordinate function (see [Ar78], section 43, or [We77], Extension Theorem, lecture 5.)

Define $\Sigma = \tilde{\Sigma} \cap \{H = H_0\}$. It is a standard fact (true for periodic orbits of general flows) that the Hamiltonian flow h^t admits a Poincaré return map $\mathcal{R}$, defined on Σ around $\boldsymbol{z}_0$, by $\mathcal{R}(\boldsymbol{z}) = h^{t(\boldsymbol{z})}(\boldsymbol{z})$, where $t(\boldsymbol{z})$ is the first return time of $\boldsymbol{z}$ to Σ under the flow (see Hirsh & Smale (1974), Chapter 13).

We claim that $\mathcal{R}$ is symplectic, with the symplectic structure induced by Ω_0 on Σ.

Since $\tilde{\Sigma}$ is transverse to γ, we may assume that:

$$\dot{q}_0 = \frac{\partial H}{\partial p_0} \neq 0$$

Fig. 60. 2.

on $\tilde{\Sigma}$. Hence, by the Implicit Function Theorem, the equation

$$H(0, q_1 \ldots, q_n, p_0, \ldots, p_n) = H_0$$

implies that p_0 is a function of $(q_1, \ldots, q_n, p_1, \ldots, p_n)$. This makes the latter variables a system of local coordinates for Σ, and since $dq_0 = 0$ on Σ, the restriction of Ω_0 is in fact

$$\omega = \Omega_0\big|_\Sigma = \sum_{k=1}^{n} dq_k \wedge dp_k.$$

To prove that $\mathcal{R}$ is symplectic, remember that, by (60.9) , for any closed curve in Σ, or more generally for any closed 1–chain c in Σ,

$$\int_{\mathcal{R}c} \boldsymbol{p}d\boldsymbol{q} - Hdt = \int_c \boldsymbol{p}d\boldsymbol{q} - Hdt$$

since c and $\mathcal{R}c$ are on the same trajectory tube. Here $\mathcal{R}c$ represent the chain in $\mathbb{R}^{2n+2} \times \mathbb{R}$ given by $(\mathcal{R}(c(s)), t^{c(s)})$.This equality implies that the function $S(\boldsymbol{z}) = \int_{\boldsymbol{z}_0}^{\boldsymbol{z}} \mathcal{R}^*(\boldsymbol{p}d\boldsymbol{q} - Hdt) - (\boldsymbol{p}d\boldsymbol{q} - Hdt)$ is well defined. But, on Σ, the differential of the form inside this integral is $\mathcal{R}^*\omega - \omega$, since both dq_0 and dH are zero there. Hence $\mathcal{R}^*\omega - \omega = d^2S = 0$, i.e., $\mathcal{R}$ is symplectic.

Remark SGleginv is 60.1,Theorem SGthmcanva is 56.1Theorem SGhamexactsymp is 60.7, Formula SGintinv is (60.9) , Formula SGst is (60.10) , Exercise SGexoexactsympcurve is 59.7, Exercise SGexoxhev is 60.13, Exercise SGexoisotopy is 60.4, Formula SGformhomotopy is (60.6) , Exo SGexolagsym is 56.6.

Appendix 2 or TOPO

SOME TOPOLOGICAL TOOLS

December 30, 1999

In order to estimate the minimum number of periodic orbits a symplectic twist map or a Hamiltonian system may have, we need an estimate on the minimum number of critical points for the energy function of the corresponding variational problem. Estimating the number of critical points of functions on compact manifolds is the jurisdiction of Morse Theory and Lyusternick-Schnirelman Theory. Given the gradient flow of a real valued function f on a compact manifold M, Morse Theory rebuilds M from the unstable manifolds of the critical points of f. The combinatorial data of this construction gives a relationship between the set of critical points and the topology of M, in the guise of its homology. Unfortunately, the space on which the energy function W is defined is not compact. However, it usually is a vector bundle over a compact manifold M, and reasonably natural boundary conditions on the map or Hamiltonian system translates into some conditions of "asymptotic hyperbolicity" for W. This is a situation where Conley's theory, which studies the relationship between the recurrent dynamics of general flows and the topology of (pieces of) their phase spaces was brought to bear with great success.

For the reader who has no background in Algebraic Topology, we start in Section 61 by outlining an easy way to compute the homology of a manifold by decomposing it into cells. We then illustrate Morse theory by considering the cells given by the unstable manifolds of critical points of a real valued function on the manifold. We hope that this will give such a reader at least a flavor of the rest of this chapter. Starting Section 63 , we assume familiarity with algebraic topology. We give the basic definitions of Conley's theory and state results on estimates of number of critical points in isolated invariant sets for gradient flows. In Section 64, we prove these results. In Section 65, we apply these results to functions on vector bundles whose gradient flow are asymptotically hyperbolic.

61.* Hands On Introduction To Homology Theory

To a manifold, or to certain subspaces of it, we want to associate some algebraic objects called homology groups that are invariant under homeomorphisms or other natural topological deformations. Usually, the best way to calculate these groups (but not the best way to show their invariance properties), is to decompose the spaces studied into well understood pieces, and then define the groups from the combinatorial data describing how these pieces fit together. In this introduction we decompose spaces into *cells*, which are discs of different dimensions, and show how to compute cellular homology.

A. Finite Cell Complexes

Given a topological space X (e.g. a differentiable manifold) we can construct a new one by *attaching a cell* of dimension n. This is done by choosing an *attaching map* f from the bounding sphere $\mathbb{S}^{n-1}$ of the *cell* D^n (a disk of dimension n) to X. The new space, denoted by $X \cup_f D^n$ is given by the union of X and D^n where each point of ∂D^n is identified with its image by f in X. The topology on $X \cup_f D^n$ is that of the quotient $X \cup D^n / \{x \sim f(x)\}$.

Examples 61.1 One can construct the sphere $\mathbb{S}^2$ by attaching the disc D^2 to a point p. The space $X = \{p\}$ is a manifold of dimension 0, and the attaching map f sends each point of the boundary circle of D^2 to p. One can also construct a sphere by attaching a disk to another one (what is the attaching map?). These constructions have obvious generalization to higher dimensions.

A *cellular space* is a space built by attaching a finite number of cells (successively), starting from a finite number of points (cells of dimension 0). If in this process each cell is attached to cells of lower dimensions, the space obtained is called a *finite cell complex* or *CW complex*. The union of all cells of dimension less than k in a finite cell complex is called the *k–skeleton*. Thus the $k+1$–skeleton is built by attaching cells of dimension $k+1$ to the k–skeleton. The dimension of the cell of maximum dimension in a cellular space X (and hence of a CW complex) is called the *dimension* of X, denoted by the usual $dim X$.

Examples 61.2 The torus can be decomposed (not in a unique way!) into a finite cell complex: its 0–skeleton is the point z. To get the 1–skeleton we attach both extremities of the "meridian" a and the "equator" b to z. The attaching maps send the boundaries -1 and 1 of the 1–cells $a \cong [-1, 1] \cong b$ to the point z. Finally, the 2–skeleton is obtained by attaching the disk D (stretched to a square) to the 1–skeleton as indicated by the "flat" picture of the torus. Note that the 1–skeleton looks like a "bouquet of two circles".

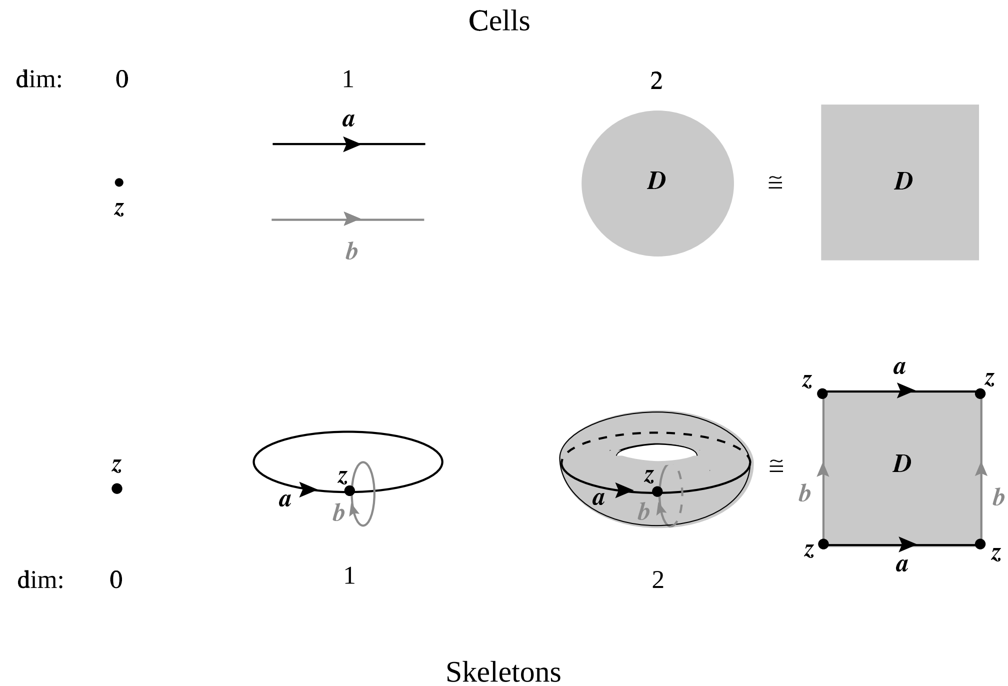


Fig. 61. 2. The torus $\mathbb{T}^2$ as a finite CW complex.

One can generalize this construction to surfaces of any genus g (spheres with g handles) by gluing a 2 cell to a polygon with $4g$ sides and identifying all vertices to a single point, and edges two by two as indicated by their name and orientations on the following figure (g is 2 here):

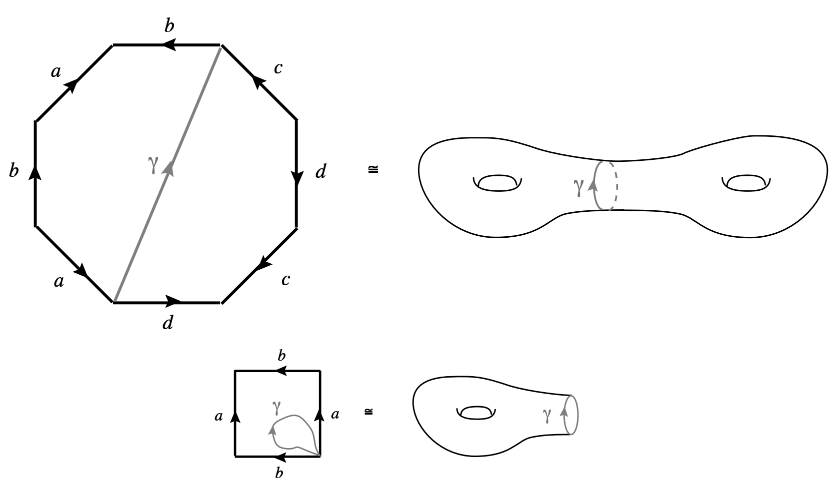


Fig. 61. 3. The double torus (surface of genus 2) as a finite CW complex. Identify edges according to their names and orientations, and identify all vertices to one point. When cutting the octagon in half through the curve γ we obtain two *handles*, which are tori with a disk (bounded by the curve γ) removed in each.

More generally, we will show in the next section that any compact manifold is homeomorphic to a finite cell complex.

Exercise 61.3 Decompose $\mathbb{S}^n, \mathbb{T}^n, \mathbb{RP}^n$ and the Klein bottle into finite cell complexes. Remember that $\mathbb{RP}^n$ can be defined as $\mathbb{D}^n/\sim$, where the relation $\sim$ identifies any two antipodal points on the boundary of the n–disk $\mathbb{D}^n$. The *Klein bottle* is $[-1,1]^2/\sim$ where $(1,y) \sim (-1,-y)$ and $(x,1) \sim (x,-1)$.

B. Cellular Homology

Bouquets of spheres. When we "crush" the (k-2)–skeleton X_{k-2} of a finite cell complex X to a point inside the (k-1)–skeleton X_{k-1}, the boundary of each (k-1)-cell crushes to that point. Hence each (k-1)-cell of X_{k-1} becomes a (k-1)-sphere in X_{k-1}/X_{k-2}. All these spheres meet at exactly one point, where the crushed X_{k-2} collapsed: we say that X_{k-1}/X_{k-2} is a *bouquet of spheres.* The attaching map f of a k–cell to X_{k-1} gives rise to a map $\tilde{f} : \mathbb{S}^{k-1} \to X_{k-1}/X_{k-2}$, by composition with the quotient map. Hence we have a map $\tilde{f}$ from a sphere of dimension $k-1$ to a bouquet of spheres, all of dimension $k-1$.

Digression on degree and homotopy. Any continuous map from a sphere S_1 to a sphere S_2 of same dimension comes equipped with a *degree*, which, informally, is an integer which measures the number of times S_1 "wraps around" S_2 under this map. This integer can be negative, as we keep track of orientation. Since the proper topological definition of degree requires homology (which we are in the process of defining), we restrict ourselves to differentiable maps. The degree of a differentiable map f between two manifolds of same dimension is given by:

$$deg(f) = \sum_{\boldsymbol{x} \in f^{-1}(\boldsymbol{z})} 1 \cdot (sign \det Df_{\boldsymbol{x}}) \tag{61.1}$$

where $\boldsymbol{z}$ is any regular value of f, i.e. the determinants in the above sum are not zero (by Sard's theorem, almost all values of a smooth map are regular). It turns out that the above number is independent of the (regular) point $\boldsymbol{z}$. The degree of a map is invariant under homotopy of the map. [Two continuous maps f_0 and f_1 between the manifold M and the manifold N are *homotopic* if there is a continuous map $F : [0,1] \times M \to N$ such that $F(0,\boldsymbol{z}) = f_0(\boldsymbol{z}), \quad F(1,\boldsymbol{z}) = f_1(\boldsymbol{z})$ for all $\boldsymbol{z}$ in M.]

Back to horticulture. The attaching map $\tilde{f} : \mathbb{S}^{k-1} \to X_{k-1}/X_{k-2}$ has a multiple degree: on each sphere S_i in the bouquet one can compute the oriented number of preimages under $\tilde{f}$ of a regular point as in (61.1) (without loss of generality, we can assume that $\tilde{f}$ is differentiable except at the common point of the spheres). Suppose that $\boldsymbol{c}_1^{k-1}, \ldots, \boldsymbol{c}_{N_{k-1}}^{k-1}$ denote the $(k-1)$–cells of the cell complex and $\boldsymbol{c}_1^k, \ldots, \boldsymbol{c}_{N_k}^k$ its k-cells. We now form an $N_k \times N_{k-1}$ integer matrix ∂_k whose entry $\partial_k(ij)$ is the degree of the attaching map from $\partial \boldsymbol{c}_i^k$ to the jth sphere of the bouquet, i.e. $\boldsymbol{c}_j^{k-1}/\partial \boldsymbol{c}_j^{k-1}$. The matrices ∂_k, for $k \in \{1, \ldots, dimX\}$ essentially give all the combinatorial data describing how the complex X is pieced together from our collection of cells.

Chain complexes. We now want to view the matrices ∂_k as those of linear maps between finite dimensional vector spaces, or modules. To do this, one thinks of $\boldsymbol{c}_1^k, \ldots, \boldsymbol{c}_{N_k}^k$ as the basis vectors of an abstract vector space (or free module) C_k whose elements are formal sums of the form

$$\boldsymbol{c} = \sum_1^{N_k} a_j \boldsymbol{c}_j^k,$$

where a_j is an element of some "coefficient" field (or ring) K (usually $\mathbb{Z}_2, \mathbb{Z}, \mathbb{Q}$ or $\mathbb{R}$). Hence C_k is generated by the k-cells and $dimC_k = N_k$. For convenience, we define $\partial_0 \equiv 0$ on C_0.

Lemma 61.4

$$\partial_{k-1} \circ \partial_k \equiv 0. \tag{61.2}$$

(The proof of this crucial lemma, which we will not give here (see, *eg.* Dubrovin & al. (1987)) usually uses the long exact sequence of a triple and a pair in simplicial homology). A chain of maps and vector spaces (or modules):

$$C_n \overset{\partial_n}{\to} C_{n-1} \to \ldots C_k \overset{\partial_k}{\to} C_{k-1} \to \ldots \to C_0$$

satisfying (61.2) is called a *chain complex*.

Definition 61.5 The kth *homology group* of the finite cell complex X with coefficients in a ring K is given by:

$$H_k(X;K) = Ker\ \partial_k / Im\ \partial_{k+1}.$$

where, by convention, $\partial_0 = 0 = \partial_{n+1}$

This definition makes sense since, by Lemma 61.4, $Im\ \partial_{k+1} \subset Ker\ \partial_k$.

Note that $H_k(X) = 0$ whenever $k > \dim X$ or $k < 0$, since for such k, $X_k = \emptyset$.

Example 61.6 The circle $\mathbb{S}^1$ is a CW complex: we start with a point p and attach to it an interval $I = [0,1]$: the boundary points of I become identified to p under the attaching map. Using $\mathbb{R}$ as coefficients in our chain complex, we get $C_0 = \mathbb{R}.p \cong \mathbb{R}, C_1 = \mathbb{R}.I \cong \mathbb{R}$. The map $\partial_1 = 0$: p has the two preimages $\{0\}$ and $\{1\}$ under the attaching map, but they come with opposite orientation under the orientation induced by I.

Example 61.7 Figure 61. 2 gives the generators for a chain complex for the torus: $C_0 = \mathbb{R} \cdot \boldsymbol{z}, C_1 = \mathbb{R} \cdot a \oplus \mathbb{R} \cdot b, C_2 = \mathbb{R} \cdot D$. All the boundary maps are 0 in this case: $\partial_1 a = 0$ because it geometrically yield z twice but with opposite orientation. Likewise for $\partial_1 b$. As for $\partial_2 D = a + b - a - b = 0$, again due to orientation. Hence $Ker\ \partial_k = Im\ \partial_{k+1} = C_k$ for $k = 1, 2, 3$. We have shown:

$$H_k(\mathbb{T}^2, \mathbb{R}) \cong \begin{cases} \mathbb{R} & k = 0 \\ \mathbb{R}^2 & k = 1 \\ \mathbb{R} & k = 2. \end{cases}$$

Clearly, this result remains valid if we replace $\mathbb{R}$ by any coefficient ring K.

Example 61.8 A less trivial example is given by the Klein bottle. This non orientable surface is a torus with a twist and it cannot be embedded in $\mathbb{R}^3$. We build it with the same cells $\boldsymbol{z}, a, b$ and D as the torus. The only change occurs in the definition of ∂_2: instead of gluing D to two copies of b in opposite orientation, we give them the same orientation (see Figure 61.4). As a result, the matrix of ∂_2 is now $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$. Let us use the integers $\mathbb{Z}$ as our coefficient ring. Then $Ker\ \partial_2 = \{0\}$. From this we immediately get that $H_2(Klein, \mathbb{Z}) = 0$. As in the torus, $\partial_1 = 0$ so that $Ker\,\partial_1 = C_1 = a \cdot \mathbb{Z} \oplus b \cdot \mathbb{Z}$. Since $Im\,\partial_2 = \{0\} \cdot a \oplus 2\mathbb{Z} \cdot b$, $H_1(Klein, \mathbb{Z}) \cong \mathbb{Z} \oplus \mathbb{Z}_2$. As in the case of the torus, $H_0(Klein, \mathbb{Z}) = \mathbb{Z}$ (in fact, the rank of H_0 gives the number of connected components of a manifold). Now let's reexamine the above computation with coefficients $K = \mathbb{Z}_2$ instead: the map $\partial_2 = 0$ in this case since 2=0 in this ring. Thus, in this case we are back to the same situation as with the torus: $H_0(Klein, \mathbb{Z}_2) \cong \mathbb{Z}_2$, $H_1(Klein, \mathbb{Z}_2) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$, $H_2(Klein, \mathbb{Z}_2) \cong \mathbb{Z}_2$. Finally, let's choose $K = \mathbb{R}$. Since $Ker\,\partial_2 = C_2$ in this case again, $H_2(Klein, \mathbb{R}) = \mathbb{R}$. Since $\mathbb{R}/2\mathbb{R} = \mathbb{R}/\mathbb{R} = \{0\}$, $H_1(Klein, \mathbb{R}) \cong \mathbb{R}$. As before $H_0(Klein, \mathbb{R}) \cong \mathbb{R}$.

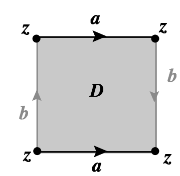


Fig. 61. 4. A cell decomposition for the Klein bottle. The only difference with that of the torus is the orientation of one of the segments b.

Some general properties and definitions related to homology. Let X be a compact manifold of dimension n. As we will see in next section, it can always be decomposed into a finite CW complex.

•$dim H_k(M, \mathbb{R}) = rank\ H_k(M, \mathbb{Z}) = b_k$ is the k^{th} *Betti number* of M.

•$\sum_{k=1}^{n} (-)^k b_k = \chi(M)$ is the *Euler characteristic* of M.

•Neither b_k nor $\chi(M)$ depend on the chain decomposition chosen for M.

•b_0 gives the number of connected component of M.

•$b_n = 1$ if M is orientable, $b_n = 0$ if M is not orientable.

Topological invariance. The importance of homology stems in great part from its invariance under topological equivalences. One topological equivalence is that of homeomorphism. A coarser equivalence (see Exercise 61.10) is that of homotopy type. Two topological spaces M and N have the same *homotopy type* if there are continuous maps $\phi : M \to N$, $\psi : N \to M$ such that $\phi \circ \psi$ and $\psi \circ \phi$ are homotopic to the Identity map of M and N respectively. In other words M can be deformed into N and vice-versa.

Theorem 61.9 *If the two manifolds M and N are homeomorphic, or have the same homotopy type, then they have same homology: $H_*(M) = H_*(N)$ (the star $*$ stands for any integer).*

Since the degree of the attaching maps are invariant under homotopy, homology is itself an invariant under homotopy equivalence (this requires a more rigorous proof, of course, see *eg.* Dubrovin & al. (1987)).

C. Cohomology

Roughly speaking, cohomology is dual to homology. For readers of this book, it might be easier to see it through differential forms, which are dual to chains of cells in the sense that the integral $< c, \omega >= \int_c \omega$ of a form ω on a chain c is a linear, real valued function in c (it is also linear in ω). The duality bracket given by integration also satisfies:

$$< \partial\, c, \omega >=< c, d\omega >$$

where d is the exterior differentiation on forms. This formal equality is a general requirement for defining cohomology. In the case of forms it is simply given by Stokes' Theorem. Finally, we can define the cochain complex

$$C_0^* \xrightarrow{d_1} C_1^* \xrightarrow{d_2} \dots \xrightarrow{d_n} C_n^*$$

where $C_k^* = \Lambda^k$ is the vector space of k–forms and d_k is exterior differentiation. As with homology, we can define the *DeRham cohomology group* as:

$$H^k(M, \mathbb{R}) = Ker\ d_{k+1}/Im\ d_k,$$

$i.e.$ this cohomology is the quotient of closed forms over exact forms. One notable difference between homology and cohomology is the direction of the arrows in the complexes that defines them. Another notable difference, which makes the use of cohomology often preferable, is the existence of a natural product operation in cohomology, called the *cup product*. In DeRham cohomology, this cup product takes the form of wedge product of the forms:

$$[\omega_1] \cup [\omega_2] = [\omega_1 \wedge \omega_2]$$

where the notation $[\omega]$ denotes the class of the closed form ω. There are many different ways to define cohomology, but it can be shown that (given some normalization requirements), they all give the same result on compact manifolds. Poincaré, for instance, introduced cohomology (not under that name) by geometrically constructing a dual complex to a triangulation (a special CW chain decomposition). In the next section, where unstable manifolds of critical points of a Morse function will provide us with a chain decomposition, the dual decomposition can be taken to be that of stable manifolds.

D*. Covering Spaces and Fundamental Group

Covering spaces. The simple notation $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is rich in geometric and algebraic meaning. The quotient map $p : \mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ is an instance of a *covering map* in that it is a local homeomorphism which is such that each point in the torus has an *evenly covered* neighbourhood U such that $p^{-1}U$ is made out of disjoint copies of U ($eg.$ take a disk of radius less than 1 U around the point z). That makes $\mathbb{R}^2$ a *covering space* of $\mathbb{T}^2$. In a covering space, the transformations that permute points in a fiber $p^{-1}(z)$ are homeomorphisms which form a group under composition called the group of *deck transformations*. For instance, $\mathbb{Z}^2$ is the group of deck transformations of the covering space $\mathbb{R}^2 \to \mathbb{T}^2$.

Lifting of curves. One can lift curves from a space M to its covering $\tilde{M}$ in a well prescribed way: if the curve γ starts at z_0 in M, choose one $\tilde{z}_0 \in p^{-1}(z_0) \subset \tilde{M}$ to start the *lift* of γ, $i.e.$ a curve $\tilde{\gamma}$ such that $p(\tilde{\gamma}) = \gamma$. Above an evenly covered neighbourhood U of z_0, there is only one way to define $\tilde{\gamma}$, since there is only one copy of U containing our choice $\tilde{z}_0$. One then proceed by continuity, covering γ with a finite number of overlapping evenly covered neighborhoods. A curve has as many distinct lifts as there are preimages of its starting point.

Classification of covering spaces for $\mathbb{T}^2$. We can construct other covering spaces of the torus, with other groups of deck transformations. For instance, the cylinder $\mathbb{R} \times \mathbb{S}^1 = \mathbb{R}^2/\mathbb{Z}$ is a covering space of the torus with deck transformations group $\mathbb{Z}$. $\mathbb{R}^2/(2\mathbb{Z} \oplus 3\mathbb{Z}) \to \mathbb{R}/\mathbb{Z}^2$ is also a covering space which is itself a torus, but "6 times as big" as the standard one it covers. It has $\mathbb{Z}_2 \oplus \mathbb{Z}_3$ as group of deck transformations, reflecting the finite number of elements a fiber $p^{-1}(z)$ has. In general, given any normal subgroup G of $\mathbb{Z}^2$, you get a covering space $\mathbb{R}^2/G \to \mathbb{T}^2$ with deck transformations group $\mathbb{Z}^2/G$. In fact these are *all* the possible covering spaces of the torus!

The fundamental group. The above classification generalizes to any manifolds, as we will see. Given a connected manifold M, we need to find a covering space which serves the role $\mathbb{R}^2$ does for $\mathbb{T}^2$. It turns out that the defining feature $\mathbb{R}^2$ has in this context is that it is connected and *simply connected*: any loop in $\mathbb{R}^2$ is homotopic to a point, or constant loop. This makes $\mathbb{R}^2$ the *universal cover* of $\mathbb{T}^2$: it is the unique (up to homeomorphism) covering space of $\mathbb{T}^2$ which is simply connected. Its uniqueness comes from a construction which works for any manifold. Choose some point z_0 in your manifold M. Declare that two curves starting at z_0 are equivalent if they have same endpoint and are homotopic. Define the covering space $\tilde{M}$ as the set of all such equivalence classes. If $[\gamma] \in \tilde{M}$ is one such equivalence class, define the covering map as $p([\gamma]) = \gamma(1)$ (its endpoint). One can indeed show that, with the appropriate topology, this is a covering space, and its deck transformations form a group called the *fundamental group* of M, denoted by $\pi_1(M, z_0)$ or $\pi_1(M)$ in short (changing the base point yields isomorphic groups). Since a deck transformation must permute points in a fiber, $\pi_1(M)$ is the group of all homotopy classes of loops based at a chosen point, with group law given by concatenation of two loops (*i.e.* follow one, then the next, which is possible since they have same endpoints). The inverse of a loop is the same loop traversed backwards. As an example, since $\mathbb{Z}^2$ is the deck transformation for the universal cover $\mathbb{R}^2$ of $\mathbb{T}^2$, we must have $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$. More generally $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$. On the other hand $\pi_1(\mathbb{S}^n) = \{0\}$, since the sphere is itself simply connected (and thus is its own universal cover).

Classification of covering spaces of any manifold M. As stated above, we can use the universal cover $\tilde{M}$ to classify all covering spaces of the (connected) manifold M: any other covering space N of M is of the form $N \cong \tilde{M}/G$ where G is a subgroup of $\pi_1(M)$. Furthermore, $G \cong \pi_1(N)$ and if G is a normal subgroup of $\pi_1(M)$, then the deck transformations of $N \to M$ form the group $\pi_1(M)/G$. Remember that G is normal if $aGa^{-1} = G$ for any $a \in \pi_1(M)$. As an example, any subgroup of $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$ is normal, since $\mathbb{Z}^2$ is abelian.

Fundamental group vs. homology. Note that $\pi_1(\mathbb{T}^2) \cong \mathbb{Z}^2 \cong H_1(\mathbb{T}^2, \mathbb{Z})$. This is not a coincidence: both groups were constructed as equivalence classes based on closed loops. In general, a theorem of Poincaré (1895) says that $H_1(M, \mathbb{Z})$ is the abelianization of $\pi_1(M)$: it is the fundamental group made commutative. The way to abelianize a group G is by taking its quotient with the subgroup $[G, G]$ of its commutators, which are of the form $xyx^{-1}y^{-1}$. Hence we can write Poincaré's theorem as:

$$H_1(M, \mathbb{Z}) \cong \pi(M)/[\pi_1(M), \pi_1(M)].$$

To see how the case $M = \mathbb{T}^2$ fits here, note that $\mathbb{Z}^2$ is already abelian. In general, $\pi_1(M)$ can be much more complicated than $H_1(M, \mathbb{Z})$. Finally, this leads us to an important case of covering space, called *universal abelian cover* of a manifold M. It is the covering $\tilde{M}/[\pi_1(M), \pi_1(M)] \to M$ which, since the subgroup $[\pi_1(M), \pi_1(M)]$ is always normal, has (abelian) deck transformation group $H_1(M, \mathbb{Z}) \cong \pi_1(M)/[\pi_1(M), \pi_1(M)]$.

Exercise 61.10 Show that the circle and the cylinder have same homotopy type but are not homeomorphic.

Exercise 61.11 Using Exercise 61.3 compute the homology of $\mathbb{S}^n, \mathbb{T}^n, \mathbb{RP}^n$.

Exercise 61.12 Convince yourself, looking at Figure 61. 3 that the fundamental group of the double torus is $< a, b, c, d; aba^{-1}b^{-1}cdc^{-1}d^{-1} >$, *i.e.* the group generated by the elements a, b, c, d together with the relation that $aba^{-1}b{-}1cdc^{-1}d^{-1} = e$, the neutral element. What is the first homology group of the double torus? Repeat the question for surfaces of genus g.

62.* Morse Theory

We now show how any compact manifold can be described as a cellular space, with cells given by the unstable manifolds of the critical points of a Morse function. This immediately yields a relationship between critical points and Homology, in the guise of the Morse Inequalities. We first define some of these terms.

Let $f : M \to \mathbb{R}$ be a differentiable function on a manifold M. A *critical point* for f is a point $\boldsymbol{z}$ at which the differential of f is zero: $df(\boldsymbol{z}) = 0$. If f is twice differentiable, the critical point $\boldsymbol{z}$ is called *nondegenerate* if

$$\det \frac{\partial^2 f(\boldsymbol{z})}{\partial x^2} \neq 0 \tag{62.1}$$

where this second derivative is taken with respect to any local coordinates x around $\boldsymbol{z}$ on M. The function f is a *Morse function* if all its critical points are nondegenerate. One can show that there are many Morse functions on any manifold. In fact Morse functions are generic in the set of twice differentiable functions. See e.g. Guillemin & Pollack (1974) , as well as Milnor (1969) .

Note that the condition (62.1) is independent of the coordinate system. Indeed, at a critical point $\boldsymbol{z}$,

$$\frac{\partial^2 f(\boldsymbol{z})}{\partial y^2} = \frac{\partial x}{\partial y}^t \frac{\partial^2 f(\boldsymbol{z})}{\partial x^2} \frac{\partial y}{\partial x}.$$

This last formula also implies that the number of negative eigenvalues of the real, symmetric matrix $\frac{\partial^2 f(z)}{\partial x^2}$ does not depend on the coordinate system chosen around the critical point $\boldsymbol{z}$. This number is called the *Morse index* of $\boldsymbol{z}$. Qualitatively, the level set portrait of a function around a nondegenerate critical point is entirely determined by the index of the critical point. Indeed:

Lemma 62.1 (Morse Lemma) *Let $\boldsymbol{z}$ be a nondegenerate critical point for a function f on a manifold of dimension n. There is a coordinate system $\boldsymbol{x}$ around $\boldsymbol{z}$ such that:*

$$f(\boldsymbol{x}) = f(\boldsymbol{z}) - x_1^2 - \ldots - x_k^2 \ + x_{k+1}^2 + \ldots + x_n^2.$$

We refer the reader to Milnor (1969) for a proof of this lemma, which generalizes the diagonalization process (Gram-Schmidt) for bilinear forms. Since the Morse Lemma clearly implies that the critical points of a Morse functions are isolated, we have:

Corollary 62.2 *A Morse function on a compact manifold has a finite number of critical points.*

The *gradient flow* of a function f is the solution flow for the O.D.E.:

$$\dot{\boldsymbol{z}} = -\nabla f(\boldsymbol{z}). \tag{62.2}$$

The *gradient* ∇f is defined here by $\langle \nabla f, . \rangle = df(.)$, where the brackets denotes some chosen Riemannian metric. The minus sign is put in (62.2) so that F decreases along the flow:

$$\frac{d}{dt} f(\boldsymbol{z}(t)) = -\left\{\nabla f(\boldsymbol{z}(t))\right\}^2 \leq 0$$

with equality occurring exactly at the critical points. The eigenvectors corresponding to the negative eigenvalues of $\frac{\partial^2 f(z)}{\partial x^2}$ span a subspace of $T_z M$ which is tangent to the unstable manifold at z of the gradient flow: that is, the $x_1, \ldots, x_k$ plane given by the Morse Lemma. We remind the reader that the *unstable manifold* of a restpoint for a flow is the manifold of points whose backward orbit is asymptotic to the restpoint. Hence *the Morse index of a nondegenerate critical point of a Morse function is the dimension of its unstable manifold.*

Remark 62.3 Note that if the metric chosen to define the gradient is the euclidean one in the Morse coordinate chart, the $(x_1, \ldots, x_k)$ plane is itself the unstable manifold of the critical point, at least in that chart. This can always be arranged, by a local perturbation of the metric, and we will assume from now on that this is the case.

The gist of Morse theory consists in studying how the topology of the *sublevel sets*:

$$M^a = \{\boldsymbol{x} \in M \mid f(\boldsymbol{x}) \leq a\}$$

changes as a varies.

Theorem 62.4 *If there is no critical points in $f^{-1}[a, b]$, then M^a and M^b are diffeomorphic. The inclusion of M^a in M^b is a deformation retraction.*

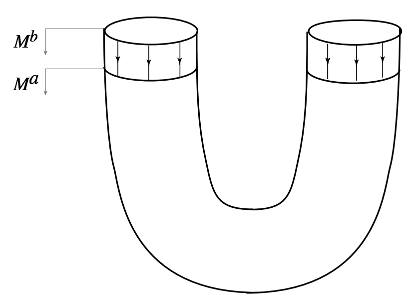


Fig. 62. 0. Deformation of a sublevel set M^b into the sublevelset M^a when there are no critical points in $f^{-1}[a, b]$. The lines with arrows represent trajectories of the gradient flow.

Proof. Deform M^b into M^a by flowing down the trajectories of the gradient flow, with appropriate speed and during an appropriate time interval. This is possible as long as there are no critical value in $[a, b]$. See Figure ??? □

Theorem 62.5 *Suppose $f^{-1}[a, b]$ is compact and has exactly one critical point in its interior, which is degenerate and of index k. Then M^b has the homotopy type of M^a with a cell of dimension k attached, namely, a ball in the unstable manifold of the critical point.*

Proof. (sketch) Let $\boldsymbol{z}$ be the critical point, $c = f(\boldsymbol{z})$ and $\epsilon > 0$ be a small real number. By the previous theorem, $M^{c+\epsilon}$ has the same homotopy type as M^b and likewise for $M^{c-\epsilon}$ and M^a. Hence, we just have to show that $M^{c+\epsilon}$ has the homotopy type of $M^{c-\epsilon}$ with a cell attached.

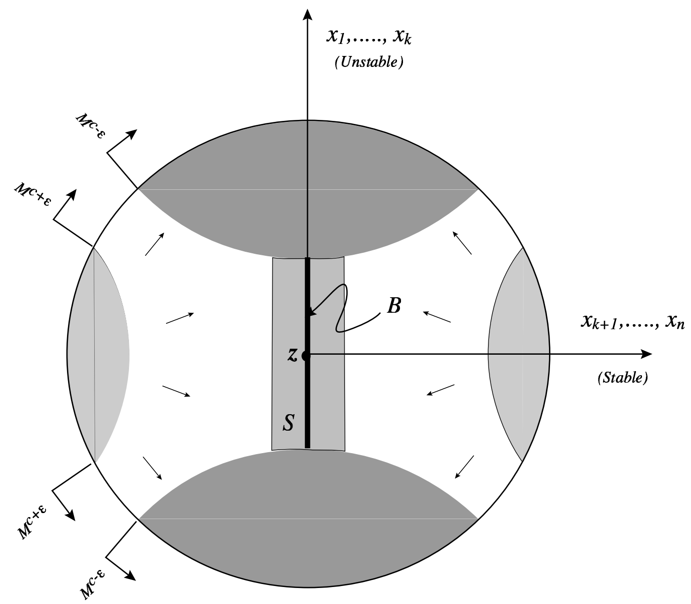


Fig. 62. 0. A neighborhood of a Morse critical point $\boldsymbol{z}$. A suitable parameterization of the flow retracts $M^{c+\epsilon}$ onto $M^{c-\epsilon} \cup S$, which itself can be deformed into $M^{c-\epsilon} \cup B$.

We have represented in Figure 62. 0 the sets $M^{c\pm\epsilon}$ within a Morse neighborhood. The drawing makes it intuitively clear that some reparameterization of the gradient flow (which we have represented by some arrows) will collapse $M^{c+\epsilon}$ into $M^{c-\epsilon} \cup S$. But the set S is given by:

$$S = \{f \leq c + \epsilon, x_1^2 + \ldots + x_k^2 \leq \delta\},$$

which can obviously be deformed into:

$$B = \{f \leq c + \epsilon, x_1 = \ldots = x_k = 0\},$$

that is, a ball in the unstable manifold of $\boldsymbol{z}$. In other words,

$$M^{c+\epsilon} \simeq M^{c-\epsilon} \cup B.$$

□

Any cellular space X is homotopically equivalent to a finite cell complex Y, where X and Y have the same number of cells in each dimension (one deforms each of the attaching maps defining X into one that attaches to cells of lower dimensions, see Dubrovin & al. (1987) , Section 4). This and the previous theorem yield:

Corollary 62.6 *Any sublevel set M^a of a Morse function on a compact manifold M has the homotopy type of a finite CW complex, whose cells correspond to the unstable manifolds of the critical points.*

Hence, since there always is a Morse function on any given manifold,

Corollary 62.7 *Any compact manifold has the homotopy type of a finite CW complex, with a cell of dimension k for each critical point of index k.*

Corollary 62.8 (Morse inequalities) *Given any Morse function f on a compact manifold M, the homology of M is generated by a finite complex $\{C_k, \partial_k\}_{\{1,\ldots,dimM\}}$ whose generators correspond to the critical points of index k of f. In particular, if $c_k = dimC_k$ is the number of critical points of index k,*

$$c_k \geq b_k = rankH_k(M, \mathbb{Z}) \tag{62.3}$$

and, better:

$$c_k - c_{k-1} + \ldots \pm c_0 \geq b_k - b_{k-1} + \ldots \pm b_0, \tag{62.4}$$

with equality holding for $k = n$.

Proof. The first statement in the theorem is somewhat of a tautology for us, since we have "defined" the homology of M as the cellular homology of any cellular complex representing M. Formula (62.3) is then trivial, since

$$H_k(M) = Ker\ \partial_k / Im\ \partial_{k+1},$$

and $Ker\ \partial_k$ is a subspace of C_k.The inequalities (62.4) are a consequence of (62.3) and their proof, left to the reader, only involves linear algebra. □

Remark 62.9 One can give a nice geometric interpretation of the maps ∂_k in the context of Morse theory. Assume that the gradient flow ϕ^t of our chosen Morse function is *Morse-Smale*, i.e. that for any given pair of critical points $\boldsymbol{x}, \boldsymbol{z}$, their respective stable and unstable manifold meet transversally. This is again a generic situation, which has the following implications: the set

$$M(\boldsymbol{x}, \boldsymbol{z}) = W^u(\boldsymbol{x}) \cap W^s(\boldsymbol{z}),$$

which is the union of all orbits connecting $\boldsymbol{x}$ and $\boldsymbol{z}$, is a manifold and

$$dimM(\boldsymbol{x}, \boldsymbol{z}) = index(\boldsymbol{x}) - index(\boldsymbol{z}).$$

In particular, if $index(\boldsymbol{x}) - index(\boldsymbol{z}) = 1$, $M(\boldsymbol{x}, \boldsymbol{z})$ is a one dimensional manifold made of a finite number of arcs that one can count, with $\pm$ according to a certain rule of intersection. This intersection number $m(\boldsymbol{x}, \boldsymbol{z})$ gives the coefficient in the generator $\boldsymbol{z}$ of $\partial(\boldsymbol{x})$, i.e.

$$\partial_k \boldsymbol{x} = \sum_{\boldsymbol{z} \in C_{k-1}} m(\boldsymbol{x}, \boldsymbol{z}).\boldsymbol{z}.$$

One can also define cohomology in this fashion: just take the same complex, but defined for the function $-f$. What was stable becomes unstable manifold and C_k becomes C_{n-k}. This not only gives us a geometric way to see cohomology, but a trivial proof of Poincaré's duality theorem:

$$H^{n-k}(M, \mathbb{R}) \cong H_k(M, \mathbb{R})$$

. For more details on this chain complex, which is sometimes called the Witten complex but dates back to J. Milnor's book on cobordism, see e.g. Salamon (1990) . For a proof of Poincaré's duality using the Morse complex, see Dubrovin & al. (1987) .

63. Controlling The Topology Of Invariant Sets

The relationship revealed by Morse between the critical point data of a function and the topology of the underlying manifold has a very wide generalization in the theory of Conley, which brings about a similar relationship for general continuous flows on locally compact topological spaces. We will outline this theory in Section 51.C. For now, we make a small step toward this generalization.

Here, and for the rest of this chapter, the cohomology used is the Čech cohomology with coefficients in $\mathbb{R}$. We do not need to define this cohomology here: it is enough to state that it is well defined not only on manifolds but on their compact subsets as well. Furthermore it is continuous for the Hausdorff topology on compact subsets. Otherwise, it satisfies all the usual axioms and rules of cohomology and coincides with other cohomologies on compact manifolds.

Consider a compact set I which is invariant under the gradient flow of a function W on some finite dimensional manifold. If W is a Morse function, then necessarily I is made of critical points and the intersections of all their stable and unstable manifolds (prove it as an exercise!). Exactly as we did for manifolds, consider the Floer-Witten chain complex, generated by the critical points and with boundary maps given by the stable-unstable manifolds intersection data. It turns out (see the proof in Floer (1989) , and also Salamon (1990)) that this complex gives the (co)homology not of I, but of its Conley index, a topological/dynamical invariant of I that we define below. In certain cases, as in what follows, one can evaluate the Conley index and hence give lower estimates on the number of critical points. We use these results in Section 65 to estimate the number of critical points of functions on vector bundles.

Definition 63.1 Let M be a finite dimensional manifold. A compact neighborhood B in M is called an *isolating block* for a (continuous) flow ϕ^t if points on the boundary ∂B of B immediately leave B under the flow, in positive or negative time:

$$z \in \partial B \Rightarrow \phi^{(0,\epsilon)} \subset B^c \quad \text{or} \quad \phi^{(-\epsilon,0)} \subset B^c \quad \text{for some} \quad \epsilon = \epsilon(z) > 0.$$

The *exit set* B^- of B is defined as the set of points in ∂B which immediately flow out of B in *positive* time.

Given an isolating block B for the flow ϕ^t, define $I(B)$ to be the *maximal invariant set* included in B ("maximal" is in the sense of inclusion here). Alternatively:

$$I(B) = \cap_{t \in \mathbb{R}} \phi^t(B).$$

There are two classical ways to measure the topological complexity of an invariant set $I(B)$. One is its cohomology *cohomology Conley index*:

$$h(I) = H^*(B, B^-).$$

The bigger the dimension of this vector space, the more complex the topology of I. Note that in the notation $h(I)$, we have deliberately omitted the mention of B: this is because the vector spaces $H^*(B, B^-)$ are isomorphic for all isolating block B such that $I = I(B)$ (Conley & Zehnder (1984)). Hence $h(I)$ is an invariant of the set I. In practice, the size of $h(I)$ is measured by the *sum of the Betti numbers*

$$sb(h(I)) = \sum_k dim H^k(B, B^-).$$

This again is an invariant of I. A second, somewhat rougher way to measure the complexity of an invariant set I (or any topological space which admits continuous (semi)flows and a cohomology) is the *cuplength* which is defined as:

$$cl(I) = 1 + \sup\{k \in \mathbb{N} \mid \exists \omega_1, \dots \omega_k, \quad \omega_j \in H^{n_j}(I), n_j > 1, \quad \text{and} \quad \omega_1 \cup \dots \cup \omega_k \neq 0\}$$

The following is a generalization of both Morse and Lyusternick-Schnirelman theories. It is itself the consequence of the much more general theory of Conley for (semi)flows.

Theorem 63.2 *Let I be a compact isolated invariant set for the gradient flow of a function W on some manifold. If the function is Morse, the number of critical points in I is greater or equal to $sb(h(I))$. Otherwise, the number of critical points is at least equal to $cl(I)$.*

Historically, the first time Theorem 63.2 was applied in a significant way was in the proof of the following proposition, which appeared in several pieces in Conley & Zehnder (1983) :

Proposition 63.3 *Let M be a compact manifold and W be a real valued function on $M \times \mathbb{R}^n \times \mathbb{R}^m$. Suppose that the gradient flow of W admits an isolating block B of the form $B \simeq M \times D^+ \times D^-$ with exit set $M \times D^+ \times \partial D^-$, where $D^+ \subset \mathbb{R}^n, D^- \subset \mathbb{R}^m$ are homeomorphic to the unit balls. If W is a Morse function, it has at least $sb(M)$ critical points in B. In general, W has at least $cl(M)$ critical points.*

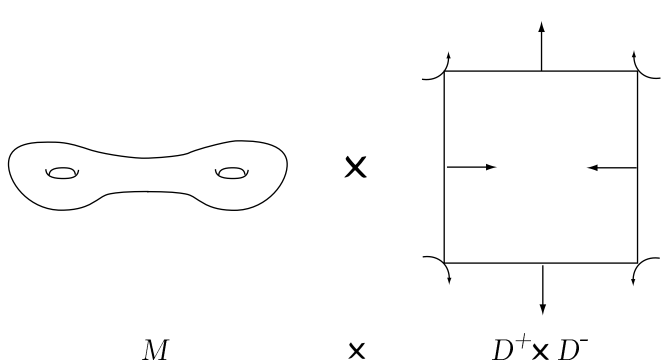


Fig. 63. 0. The isolating neighborhood in Proposition 63.3.

Conley and Zehnder applied this theorem in the case $M = \mathbb{T}^n$, where $sb(M) = 2^n$, and $cl(M) = n+1$, which gives a measure of the crudeness of the cuplength as compared to the sum of the Betti numbers. In the following section we will give a proof of Theorem 63.2 (we will only sketch the sum of betti number estimate, but give a complete proof of the cuplength estimate) as well as of Proposition 63.3.

64. Topological Proofs

The following lemma gives a situation where one can get a handle on the topology of an invariant set I. It is central to the proofs of several topological results we will use, including Proposition 63.3.

Lemma 64.1 (**Floer**) *Let B be an isolating block for a flow ϕ^t on a finite dimensional manifold, and I be its maximal invariant set. Suppose that there is a retraction $\alpha : B \to P$, where P is some compact subset of B. If there is a class $u \in H^*(B, B^-)$ such that :*

$$v \mapsto u \cup \alpha^*(v) : H^*(P) \to H^*(B, B^-)$$

is an isomorphism, then

$${\alpha_I}^* : H^*(P) \to H^*(I)$$

is injective, where α_I denotes the restriction of α to I.

(If $N \subset M$ are two topological spaces and $i : N \to M$ is the inclusion map, a *retraction* is a map $r : M \to N$ such that $r \circ i = Id_N$, that is r restricts to Id on N). For a proof of Lemma 64.1, see Section 65.B.

Corollary 64.2 *Let B, I, P be as in Lemma 64.1, and let the flow ϕ^t in that lemma be the gradient of some function W. Then the number of critical points of W is at least $cl(P)$.*

Proof. If $H^*(P) \to H^*(I)$ is injective, $cl(I) \geq cl(P)$ and the Corollary is an immediate consequence of Proposition 63.2. □

A. Proof of the Cuplength Estimate in Theorem 63.2

Conley & Zehnder (1983) prove a cuplength estimate (their Theorem 5) that is valid for a compact invariant set I of a general flow ϕ^t. We follow their proof. Define a *Morse decomposition* for I to be a finite collection $\{M_p\}_{p\in P}$ of disjoint compact and invariant subsets of I, which can be ordered in such a way that any x not in $\cup_{p\in P} M_p$ is α-asymptotic to an M_j and ω-asymptotic to an M_i, with $i < j$ (x is *α-asymptotic* (resp. *ω-asymptotic*) to M_j if $\lim_{t\to -\infty\ (+\infty)} \phi^t(x) \in M_j$). One can show that a compact invariant set always has such a Morse decomposition. We now state Theorem 5 of Conley & Zehnder (1983) :

Theorem 64.6 *Let I be any compact invariant set for a continuous flow, and let $\{M_P\}_{p\in P}$ be a Morse decomposition for I. Then*

$$cl(I) \le \sum_{p \in P} cl(M_p). \tag{64.1}$$

The relevant example for us is when ϕ^t is the gradient flow of a function with a finite number of (not necessarily nondegenerate) critical points on a compact invariant set I: it easy to check that these critical points form a Morse decomposition. Since an isolated point has trivial cohomology, $cl(M_p) = 1$ for each p in this example, and we have proven the cuplength estimate in Theorem 63.2. The case when the critical points are not isolated is trivial in that theorem: $cl(I) < \infty$ is always true... We now prove Theorem 64.6.

Proof. Note that if $(M_1, \ldots, M_k)$ is a Morse decomposition, then $(M_{1,\ldots,k-1}, M_k)$ is also a Morse decomposition, where $M_{1,\ldots,k-1}$ is formed by the union of $M_1 \cup \ldots \cup M_{k-1}$ and of all the connecting orbits between these sets. Hence, by induction, we only need to consider the case where $k = 2$, and (M_1, M_2) is a Morse decomposition for I. From the definition of a Morse decomposition, we can deduce the existence of two compact neighborhoods I_1 of M_1 and I_2 of M_2 in I with $I_1 \cup I_2 = I$ and such that $M_1 = \cap_{t>0}\phi^t(I_1)$ and $M_2 = \cap_{t>0}\phi^t(I_2)$. In particular, by continuity of the Čech cohomology $H^*(I_j) = H^*(M_j), j = 1, 2$. Thus the proof of (64.1) reduces to that of the inequality $cl(I_1) + cl(I_2) \ge cl(I)$ whenever $I_1 \cup I_2 = I$ are three compact sets. The next lemma is devoid of dynamics:

Lemma 64.8 *Let $I_1 \cup I_2 \subset I$ be three compact sets. If $i_1 : I_1 \to I$, $i_2 : I_2 \to I$ and $i : I_1 \cup I_2 \to I$ are the inclusion maps, then, for any $\alpha, \beta \in H^*(I)$,*

$$i_1^*\alpha = 0 \quad \text{and} \quad i_2^*\beta = 0 \Rightarrow i^*(\alpha \cup \beta) = 0.$$

Proof. We chase the diagram:

$$\begin{array}{ccccc}
H^*(I, I_1) & \otimes & H^*(I, I_2) & \overset{\cup}{\to} & H^*(I, I_1 \cup I_2) \\
\downarrow j_1^* & & \downarrow j_2^* & & \downarrow j^* \\
H^*(I) & \otimes & H^*(I) & \overset{\cup}{\to} & H^*(I) \\
\downarrow i_1^* & & \downarrow i_2^* & & \downarrow i^* \\
H^*(I_1) & \otimes & H^*(I_2) & \overset{\cup}{\to} & H^*(I_1 \cup I_2).
\end{array}$$

The vertical sequences are exact sequences of pairs. Starting on the second line of the diagram with $\alpha, \beta \in H^*(I)$, suppose $i_1^*\alpha = 0 = i_2^*\beta$ then there must be $\tilde\alpha \in H^*(I, I_1)$ with $j_1^*\tilde\alpha = \alpha$, $\tilde\beta \in H^*(I, I_2)$ with $j_2^*\tilde\beta = \beta$. Now $j^*(\tilde\alpha \cup \tilde\beta) = \alpha \cup \beta$ and hence $i^*(\alpha \cup \beta) = i^* \circ j^*(\tilde\alpha \cup \tilde\beta) = 0$, by exactness. □

To finish the proof of Theorem 64.6, let $\alpha_1, \ldots, \alpha_l$ be in $H^*(I)$ and $\alpha_1 \cup \ldots \cup \alpha_l \ne 0$. Let this product be maximum, so that $cl(I) = l + 1$. Order the α's in such a way that $\alpha_1 \cup \ldots \cup \alpha_r$ is the longest product not in the kernel of i_1^*. In particular $cl(I_1) \ge r + 1$ and $i_1^*(\alpha_1 \cup \ldots \cup \alpha_r \cup \alpha_{r+1}) = 0$. Lemma 64.8 forces $i_2^*(\alpha_{r+1} \cup \ldots \cup \alpha_l) \ne 0$ (i^* is one-to-one here, since $I_1 \cup I_2 = I$). Thus $cl(I_2) \ge l - (r + 1) + 1 = l - r$, and $cl(I_1) + cl(I_2) \ge l + 1 = cl(I)$. □

B. Proof of Lemma 64.1

In this subsection, we prove Lemma 64.1 that we restate here:

Lemma 64.1 **(Floer)** *Let B be an isolating block for a flow ϕ^t on a finite dimensional manifold, and I be its maximal invariant set. Suppose that there is a retraction $\alpha : B \to P$, where P is some compact subset of B. If there is a class $u \in H^*(B, B^-)$ such that :*

$$v \mapsto u \cup \alpha^*(v) : H^*(P) \to H^*(B, B^-)$$

is an isomorphism, then

$$\alpha_I{}^* : H^*(P) \to H^*(I)$$

is injective, where α_I denotes the restriction of α to I.

Proof. Define $B^\infty = \cap_{t>0}\phi^t B$, the set of points that stay in B for all negative time.

Lemma 64.9 *1) $H^*(B, B^\infty \cup B^-) = 0$*
2) $l^* : H^*(B^\infty) \to H^*(I(B))$ is an isomorphism, where $l : I(B) \to B^\infty$ is the inclusion.

Before proving this lemma, we use it to finish the proof of Lemma 64.1. Consider the diagram:

$$\begin{array}{ccccccc}
H^*(B,B^-) & \otimes & H^*(B,B^\infty) & \overset{\cup}{\to} & H^*(B,B^\infty \cup B^-) & = 0 \\
\downarrow Id & & \downarrow j^* & & \downarrow k^* & \\
H^*(B,B^-) & \otimes & H^*(B) & \overset{\cup}{\to} & H^*(B,B^-) & \\
 & & \downarrow i^* & & & \\
 & & H^*(B^\infty) & \overset{l^*}{\underset{\to}{\sim}} & H^*(I) &
\end{array}$$

where all vertical maps are induced by inclusions, and the two first horizontal maps are given by Künneth Formula. Suppose $\alpha_I{}^*v = 0$ for some $v \in H^*(P)$. Since l^* is an isomorphism and $\alpha_I = \alpha_{B^\infty} \circ l$, $0 = \alpha_I{}^* = l^*(\alpha_{B^\infty})^*v \Rightarrow (\alpha_{B^\infty})^*v = 0$. Since $\alpha_{B^\infty} = \alpha \circ i$, $0 = \alpha_{B^\infty}{}^*v = i^*\alpha^*v$. The middle, vertical sequence is the exact sequence of a pair. Hence there is a $w \in H^*(B, B^\infty)$ such that $j^*w = \alpha^*v$. But $u \cup \alpha^*v = k^*(u \cup w) = k^*(0) = 0$. The hypothesis of Lemma 64.1 forces $v = 0$. □

Proof of Lemma 64.9 Let $B^t = \phi^t(B)$ and $B^\infty = \cap_{t>0}B^t$ as before. Note in particular that, in the Hausdorff topology, $\lim_{t\to\infty} B^t = B^\infty$, and $\lim_{t\to 0} B^t = B$. To the triple of spaces $(B, B^- \cup B^t, B^-)$ corresponds the exact sequence:

$$\ldots \overset{\delta^*}{\to} H^*(B, B^t \cup B^-) \to H^*(B, B^-) \overset{i^*}{\to} H^*(B^t \cup B^-, B^-) \overset{\delta^*}{\to} H^{*-1}(B, B^t \cup B^-) \ldots,$$

(see *eg.* Dubrovin & al. (1987)). We now show that i^* is an isomorphism. Consider the diagram:

$$\begin{array}{ccc}
 & & (B^t \cup B^-, B^-) \\
 & \overset{i_1}{\nearrow} & \downarrow i \\
(B^t, B^- \cap B^t) & \overset{i_2}{\longrightarrow} & (B, B^-)
\end{array}$$

.

The excision theorem implies that i_1^* is an isomorphism, and the continuity of the Čech cohomology implies that i_2^* is an isomorphism. Since the diagram commutes, i^* must be an isomorphism. But this forces $H^*(B, B^t \cup B^-) = 0$ in the above diagram. Taking the limit of this equality as $t \to \infty$ proves 2).

Using the long exact sequence of the pair (B^∞, I), the map l^* induced by the inclusion $l : I \to B^\infty$ is an isomorphism whenever $H^*(B^\infty, I) = 0$, which we proceed to show. Note that $\phi^{-t}B^\infty \subset B^\infty$ and, by definition, $I = \cap_{t\geq 0}\phi^{-t}(B^\infty)$. Consider the maps:

$$(B^\infty, \phi^{-t}B^\infty) \overset{\phi^{-t}}{\to} (\phi^{-t}B^\infty, \phi^{-t}B^\infty) \overset{j}{\to} (B^\infty, \phi^{-t}B^\infty),$$

where j is the inclusion. The map $j \circ \phi^{-t}$ is clearly homotopic to Id, hence $H^*(B^\infty, \phi^{-t}B^\infty) \cong H^*(\phi^{-t}B^\infty, \phi^{-t}B^\infty) = 0$. Since this is true for all t, the continuity of the Čech cohomology concludes. □

C*. The Betti Number Estimate of Theorem 63.2 and Conley's Theory: a Sketch

We have proven in Theorem 64.6that, for a general function W, the number of critical points in an invariant set I for the gradient flow of W is greater than $cl(I)$. We now show that if W is a Morse function, the number of critical points in I is greater than $sb(I)$. To do so, one can either follow Floer (1989) in his generalization of the Witten complex (of unstable manifolds of critical points for gradient flows, see Remark 62.9) to invariant sets. His proof relies in part on Conley's theory. Alternatively, one can use Conley's generalized Morse inequalities that we state in this subsection.

Let I be a compact invariant set for a continuous flow ϕ^t and $(M_1, \ldots, M_k)$ be a Morse decomposition of I. Analogously to Theorem 64.6, Conley-Morse inequalities relate certain betti numbers of the Morse sets M_j to the corresponding betti numbers of I. To define the adequate betti numbers, we need to generalize the notion of isolating block to that of index pair for isolated invariant sets. A compact set I is an *isolated invariant set* if there is a neighborhood N of I such that $I = I(N)$ is the maximal invariant subset in N. An *index pair* for an isolated invariant set I is a pair of compact spaces (N_1, N_2) such that $N_1 \backslash N_2$ is a neighborhood of I and $I = I(N_1 \backslash N_2)$. This generalizes the concept of isolating block. In particular N_2 plays the role of the exit set, see Conley (1978), Conley & Zehnder (1984) . The fundamental property of these sets is that the homotopy type $[N_1/N_2, *]$ is independent of the choice of index pair for I and hence defines a topological invariant called the *Conley index* of the invariant set I. Giving less information, but easier to manipulate is the *cohomology Conley index* $H^*(N_1, N_2) = h(I)$, again an invariant of I. If $(N_1, N_2) = (B, B^-)$ for an isolating block B, this definition of $h(I)$ is the same as we have given previously. One way to encode the information given by $h(I)$ is via the coefficients of the *Poincaré polynomial*:

$$p(t, h(I)) := \sum_{j\geq 0} t^j dim H^j(N_1, N_2).$$

In Conley & Zehnder (1984) , it is proven that, given a Morse decomposition $(M_1, \ldots, M_k)$ for an invariant set I of a continuous flow ϕ^t, there is a *filtration* $N_0 \subset N_1 \subset \ldots \subset N_k$ such that (N_j, N_{j-1}) is an index pair for M_j. This is instrumental in proving the following:

Theorem 64.12 (Conley-Morse inequalities)

$$\sum_{j=1}^{k} p(t, h(M_j)) = p(t, h(I)) + (1+t)Q(t), \tag{64.2}$$

where $Q(t)$ is a polynomial with positive coefficients

This theorem is an extraordinary generalization of the classical Morse inequalities: it is valid for any *continuous* flow on a locally compact space (not necessarily a manifold!). To see that one indeed retrieves the betti number estimates of Theorem 63.2, one uses the Morse decomposition of our invariant set I given by the (isolated) critical points $z_1, \ldots, z_N$. Thanks to the Morse Lemma, it is not hard to construct an isolating block for each z_j, and show that the Conley index of z_j is a pointed sphere made by collapsing the boundary local unstable manifold of z_j to a point: take the set S in Figure 64. 1.

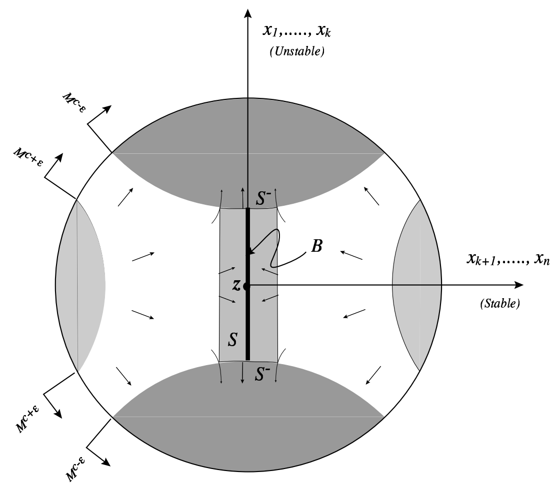


Fig. 64. 1. The index pair (S, S^-) retracts on (B, B^-), a pair made of the local unstable manifold of z and its boundary (a disk of dimension k equal to the index of the critical point z and its bounding sphere). Thus $h(z) = H^*(S, S^-) \cong H^*(B, B^-) \cong H^*(\mathbb{S}^k, *)$ which has exactly one generator in dimension k.

Hence $p(t, h(z_j)) = t^{u_j}$, where u_j is the Morse index of z_j. Now the pair $(I, \emptyset)$ is an isolating pair for I (no points exit I), and thus $p(t, h(I)) = \sum t^k dim H^k(I)$. The positivity of the coefficients of Q in (64.2) therefore insures that there are at least $dim H^k(I)$ critical points of index k.

D. Proof of Proposition 63.3

To prove this proposition, we let the manifold M play the role of P in Lemma 64.1. The retraction α of that lemma is given by the canonical projection $\alpha : B \to M$. Clearly the projection of B onto $M \times D^-$ is a deformation retract, which deforms B^- onto $M \times \partial D^-$. Hence $H^*(B, B^-) \cong H^*(M \times D^-, M \times \partial D^-)$. Now, Künneth Formula gives an isomorphism:

$$H^*(M) \otimes H^*(D^-, \partial D^-) \overset{\cup}{\xrightarrow{\sim}} H^*(M \times D^-, M \times \partial D^-)$$

where, as suggested by the notation, one gets all of the classes in the right hand side vector space as cup products of classes in the two left hand side spaces (with the appropriate identifications given by the inclusion maps). But, letting $n = dim D^-$, we have $H^*(D^-, \partial D^-) \cong H^*(\mathbb{S}^n, *)$, which has exactly one generator u in dimension n.

Hence $H^*(M) \cong H^{*+dim\ M}(B, B^-)$ and $sb(M) = sb(h(I))$ where I is the maximal invariant set in B. This and Theorem 63.2 yield the Betti number estimate. The homeomorphism $H^*(M) \cong H^*(B, B^-)$ is of the type prescribed by Lemma 64.1. This implies that the induced map $H^*(M) \to H^*(I)$ is injective and hence $cl(I) \geq cl(M)$. This fact and Theorem 63.2 give the cuplength estimate. □

E*. Floer's Theorem of Global Continuation of Hyperbolic Invariant Sets.

Floer's Lemma 64.1 is the cornerstone to the proof of the following theorem, where he makes good use of the powerful property of "invariance under continuation" of the Conley Index. This theorem illustrates the power of Conley's theory, and shows the historical root of Floer's Cohomology. Note that, in the theory of dynamical systems, the hyperbolicity of an invariant set for a dynamical system is intimately related to its persistence under *small* perturbations of the system: this relationship is the core of many theorems on structural stability. What is interesting about the following theorem (and Conley's theory in general) is that it provides situations when the persistence of an invariant set can be made global (but rough).

The notion of continuation of invariant sets makes use of the simple following fact: an index pair for a flow ϕ^t will remain an index pair for all flows that are C^0 close to ϕ^t. Two isolated invariant sets for two different flows are *related by continuation* if there is a curve of flows joining them (*i.e.* an isotopy) which can be (finitely) covered by intervals of flows having the same index pair. The following theorem (Theorem 2 in Floer (refine)) can be seen as an instance of weak, but *global*, stability of normally hyperbolic invariant sets.

Theorem 64.16 **(Floer)** *Let ϕ^t_λ be a one parameter family of flows on a C^2 manifold M. Suppose that G_0 is a compact C^2 submanifold invariant under the flow ϕ^t_0. Assume moreover that G_0 is* **normally hyperbolic**, *i.e. there is a decomposition:*

$$TM\big|_{G_0} = TG_0 \oplus E^+ \oplus E^-$$

which is invariant under the covariant linearization of the vector field V_0 corresponding to ϕ^t_0 with respect to some metric $\langle\ ,\ \rangle$, so that for some constant $m > 0$:

$$\begin{aligned} &\langle \xi, DV_0\xi\rangle \leq -m\langle \xi,\xi\rangle \text{ for } \xi \in E^- \\ &\langle \xi, DV_0\xi\rangle \geq m\langle \xi,\xi\rangle \text{ for } \xi \in E^+ \end{aligned} \tag{64.3}$$

Suppose that there is a retraction $\alpha : M \to G_0$ and that there is a family G_λ of invariant sets for ϕ^t_λ which are related by continuation to G_0 . Then the map:

$$\left(\alpha\big|_{G_\lambda}\right)^* : H^*(G_0) \to H^*(G_\lambda)$$

in Čech cohomology is injective.

In this precise sense, normally hyperbolic invariant sets continue *globally*: their topology can only get more complicated as the parameter varies away from 0.

65. Generating Phases Quadratic at Infinity

A. Generating Phases on Product Spaces

The following proposition serves a key role in various proofs in this book, as well as in symplectic topology.

Proposition 65.1 *Let M be a compact manifold, and W a real–valued function on $M \times \mathbb{R}^K$ satisfying:*

$$\lim_{\|\boldsymbol{v}\|\to\infty} \frac{1}{\|\boldsymbol{v}\|}\left(\frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{q},\boldsymbol{v}) - d\mathcal{Q}(\boldsymbol{v})\right) = 0, \tag{65.1}$$

where $\mathcal{Q}(\boldsymbol{v})$ is a nondegenerate quadratic form on $\mathbb{R}^K$. Then W has at least $cl(M)$ critical points. If W is a Morse function, then it has at least $sb(M)$ critical points.

The function W of Proposition 65.1 is a special case of a class of function called *generating phases*. We develop this notion in the next subsection.

Proof. In an appropriate orthonormal basis $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_K)$ of $\mathbb{R}^K$,

$$\mathcal{Q}(\boldsymbol{v}) = \langle A\boldsymbol{v}, \boldsymbol{v}\rangle \quad \text{with} \quad A = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_K \end{pmatrix},$$

with $a_j \neq 0$. Let $(v_1, \dots, v_K)$ be the coordinates of an element $\boldsymbol{v} \in \mathbb{R}^K$ in this basis. We claim that:

$$B(C) = \{(\boldsymbol{q},\boldsymbol{v}) \in M \times \mathbb{R}^K \mid \sup_j |v_j| \leq C\}$$

is an isolating block for the gradient flow of $-W$, when C is large enough.

To prove this, note that $B(C)$ is a compact neighborhood. Thus, in order to show that the flow exits in small positive or negative time at the boundary of $B(C)$, it suffices to check that, on each "face" $\{v_j = C\}$ of $\partial B(C)$, the dot product of ∇W with the normal vector to this face is non zero. The same argument will apply to the face $\{v_j = -C\}$. The normal unit vector pointing out at a point $\boldsymbol{z} = (\boldsymbol{q},\boldsymbol{v})$ of $\{v_j = C\}$ is $\boldsymbol{e}_j$. But:

$$\begin{aligned} \frac{\partial W}{\partial \boldsymbol{v}}(\boldsymbol{z}).\boldsymbol{e}_j &= \frac{\partial \mathcal{Q}}{\partial \boldsymbol{v}}.\boldsymbol{e}_j + \left(\frac{\partial W}{\partial \boldsymbol{v}} - \frac{\partial \mathcal{Q}}{\partial \boldsymbol{v}}\right).\boldsymbol{e}_j \\ &= C\left(a_j + \frac{1}{C}\left(\frac{\partial W}{\partial \boldsymbol{v}} - \frac{\partial \mathcal{Q}}{\partial \boldsymbol{v}}\right).\boldsymbol{e}_j\right) \end{aligned}$$

This last expression must be of the sign of a_j, for large C as the last term inside the bracket tends to zero when $C \to \infty$ ($\|\boldsymbol{v}\|$ is of the order of C.) The same proof works for the face $\{v_j = -C\}$, since the outward normal vector is $-\boldsymbol{e}_j$ on this face. We have proved that, for all C larger than some C_0, the set $B(C)$ is an isolating block. Denote by B^- the *exit set* of $B = B(C)$, *i.e.* the subset of ∂B on which points flow out in positive time. In this case B^- is the union of the faces $\{v_j = \pm C\}$ such that the corresponding eigenvalue a_j is negative (remember, we are looking at the gradient flow of $-W$).

Hence $B \cong M \times D^+ \oplus D^-$ where the disks D^+, D^- are respectively the intersections of the positive and negative eigenspaces of $\mathcal{Q}$ with the set $\{\sup_j |v_j| \leq C\}$, and the exit set is $B^- \cong M \times D^+ \oplus \partial D^-$. We are exactly in the situation of Proposition 63.3 which gives us the appropriate estimates for the number of critical points inside B. □

B. Generating Phases on Vector Bundles

Proposition 65.1 is a cornerstone in the theory of generating phases. We now develop this theory a little and prove a generalization of Proposition 65.1 for functions on non trivial bundles which we will need. In Chapter 9, we will show how this theory gives an approach to symplectic topology.

Definition 65.2 A *generating phase* is a function

$$W : E \to \mathbb{R}$$

where E is the total space of a vector bundle $E \to M$ and M a manifold.

If moreover W satisfies:

$$\lim_{\|\boldsymbol{v}\|\to\infty} \frac{1}{\|\boldsymbol{v}\|}\left(\frac{\partial}{\partial \boldsymbol{v}}(W - \mathcal{Q})\right) = 0, \qquad (65.2)$$

where, for each $\boldsymbol{q}$, $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v})$ is a nondegenerate quadratic form with respect to the fiber $\boldsymbol{v}$, then we say W is a *generating phase quadratic at infinity* , abbreviated *g.p.q.i.*.

We will see in Chapter 9 that the term "generating" refers to the fact that, provided they satisfy a generic condition in their derivative, generating phases generate Lagrangian manifolds of T^*M. Generating phases are also called *generating functions* when associated to the Lagrangian manifold that they generate, or *generating phase function*. We will show in Chapter 9 that twist maps generating functions *are* generating functions in this sense. We now define some elementary operations on generating phases. These will enable us to extend Proposition 65.1 to cover general g.p.q.i.'s. These operations are specially important in symplectic topology in that they enable one to define symplectic invariants of Lagrangian manifolds (capacities) as minimax values of their generating functions (see Viterbo (1992) and Siburg (1995)).

Definition 65.3 Let $W_1 : E_1 \to \mathbb{R}$, and $W_2 : E_2 \to \mathbb{R}$ be two generating phases. We say that W_1 and W_2 are *equivalent* if there is a fiber preserving diffeomorphism $\Phi : E_1 \to E_2$ such that:

$$W_2 \circ \Phi = W_1 + cst.$$

Definition 65.4 Let $W_1 : E_1 \to \mathbb{R}$ be a g.p.q.i. and $f : E_2 \to \mathbb{R}$ a nondegenerate quadratic form in the fibers of E_2. The function $W_2 : E_1 \oplus E_2 \to \mathbb{R}$ defined by:

$$W_2(\boldsymbol{q}, \boldsymbol{v}_1, \boldsymbol{v}_2) = W_1(\boldsymbol{q}, \boldsymbol{v}_1) + f(\boldsymbol{q}, \boldsymbol{v}_2)$$

is called a *stabilization* of W_1.

Proposition 65.5 *If the generating phase W_1 is equivalent to W_2, or is a stabilization of W_2 (or both) then critical points of W_1 are mapped bijectively into those of W_2 and the set of critical values are the same, up to a shift by a constant.*

Proof. Let $W_1 \circ \Phi = W_2 + C$ as in Definition 65.2. Then,

$$dW_1 = \Phi^* dW_2$$

Hence the set of critical points of W_1 is sent bijectively to that of W_2 by Φ. There is a constant discrepancy of C between critical values of W_1 and W_2 in this case .

Now let

$$W_2(\boldsymbol{q}, \boldsymbol{v}_1, \boldsymbol{v}_2) = W_1(\boldsymbol{q}, \boldsymbol{v}_1) + f(\boldsymbol{q}, \boldsymbol{v}_2)$$

be as in Definition 65.3. Critical points of a generating phases W satisfy, in particular, $\partial W/\partial \boldsymbol{v} = 0$. But here,

$$\frac{\partial W_2}{\partial \boldsymbol{v}} = (\partial W_1 \boldsymbol{v}_1, \partial f \boldsymbol{v}_2) = 0 \Rightarrow \boldsymbol{v}_2 = 0$$

and since any point $(\boldsymbol{q}, 0)$ of E_2 is critical for f, the critical points of W_2 correspond exactly to those of W_1. It is easy to see that the critical values of W_1 and W_2 are the same at the corresponding critical points. □

Proposition 65.6 *Let M be a compact manifold and $W : E \to \mathbb{R}$ be a g.p.q.i. on a fiber bundle $E \to M$. Then W has at least $cl(M)$ critical points. If W is a Morse function, then it has at least $sb(M)$ critical points.*

Proof. It is a corollary of Proposition 65.1 and of the following:

Lemma 65.7 *Let $W : E \to \mathbb{R}$ be a g.p.q.i. Then it is equivalent, after stabilization, to a g.p.q.i. $\overline{W} : M \times \mathbb{R}^K \to \mathbb{R}$ whose quadratic part $\overline{\mathcal{Q}}$ is independent of the base point.*

Proof. (We follow Theret (1999)) There exists a fiber bundle F such that $E \oplus F$ is trivial (*eg.* take F to be the dual of E, see Klingenberg (1982)). Stabilize W by endowing F with a nondegenerate quadratic form $\mathcal{Q}_2$. Since $E \oplus F$ is trivial, there is a fiber bundle diffeomorphism $\Phi : E \oplus F \to M \times \mathbb{R}^K$. A fiber bundle diffeomorphism being linear in each fiber, $(W \oplus \mathcal{Q}_2) \circ \Phi^{-1}$ is a g.p.q.i. on $M \times \mathbb{R}^K$.

We now show that any g.p.q.i. $W(\boldsymbol{q}, \boldsymbol{v})$ on a trivial bundle $M \times \mathbb{R}^K$ is equivalent to one with a quadratic part which is independent of the base point $\boldsymbol{q}$. Let $\mathcal{Q}$ be the quadratic part of W and write $\mathcal{Q}(\boldsymbol{q}, \boldsymbol{v}) = \langle A(\boldsymbol{q})\boldsymbol{v}, \boldsymbol{v}\rangle$, where $\langle\, , \rangle$ denotes the dot product on $\mathbb{R}^K$. Let $E_{\boldsymbol{q}}^+ \oplus E_{\boldsymbol{q}}^- = E_{\boldsymbol{q}}$ be the decomposition of $E_{\boldsymbol{q}}$ into the positive and negative eigenspaces of $A(\boldsymbol{q})$. If the fiber bundles E^+ and E^- were trivial, the Gram-Schmidt diagonalization process would make $\mathcal{Q}$ equivalent to a constant quadratic form with matrix $\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$ on $E^+ \oplus E^-$, and the resulting fiber bundle diffeomorphism would make W equivalent to a g.p.q.i. such as we advertised. To arrive to this situation, stabilize $\mathcal{Q}\big|_{E^+}$ (*resp.* $\mathcal{Q}\big|_{E^-}$) to a positive definite $\mathcal{Q}^+$ (*resp.* negative definite $\mathcal{Q}^-$) on a *trivial* bundle $\mathcal{E}^+$ (*resp.* $\mathcal{E}^-$). □

Remark 65.8 Our definition of g.p.q.i. is more general than the one commonly found in the (french) literature (*i.e.* Sikorav (1986), Laudenbach & Sikorav (1985), Chaperon (1989), Theret (1999), Viterbo (1992)). Usually one asks that *W be equal* to its quadratic part $\mathcal{Q}$ outside of a compact set. One can show (see Theret (1999)) that if $W - \mathcal{Q}$ is bounded outside of a compact set, then W is equivalent, after stabilization, to such a g.p.q.i.. It is not clear to us that the same would hold with our more general asymptotic condition. In that sense, Proposition 65.6is stronger of its kind than any we know of in the literature.

Proposition gpqi or TOPOpropgpqi is 65.6, Proposition TOPOproptrivialgpqi is 65.1Theorem floerthm or TOPOthmfloer is 64.16, Proposition TOPOpropcz is 63.3, Section TOPOsectionproofs is 64, Section TOPO-sectioninvtset is 63, Lemma TOPOlemfloer is 64.1, Theorem TOPOthmsbcl is 63.2, Section TOPOsecgpqi is 65

Bibliography

April 19 1999

This is still full of holes!

Abraham, R. and Marsden, J. (1985) Foundation of Mechanics, 3d Edition, Addison–Wesley

Aubry, S. and Le Daeron, P.Y. (1983) The discrete Frenkel-Kontorova model and its extensions. I. Exact results for ground states, Physica 8D

Aubry, S. and Abramovici, G. (1990) Chaotic trajectories in the standard map: the concept of anti–integrability, Physica D 43

S. Aubry, R. MacKay & C. Baesens (1991) Equivalence of uniform hyperbolicity for symplectic twist maps and phonon gap for the Frenkel-Kontorova models, Physica D 56

Angenent, S. (1988) The Periodic orbits of an area preserving twist map, Comm. in Math. Physics vol 115, no3

Angenent, S. (???) Monotone recurrence relations, their Birkhoff orbits and topological entropy; Ergod. Th. and Dynam. Sys. 10

Angenent, S. & Golé, C. (1991) Lamination by Ghost Circles, ETH, Zürich.

Arnaud, M.C. (1992) Sur les points fixes des difféomorphismes exacts symplectiques de $T^*\mathbb{T}^n \times \mathbb{R}^n$, C.R. Acad. Sci. Paris, t.309, Série I

Arnaud, M.C. (1992) Type des points fixes des difféomorphismes symplectiques de $T^*\mathbb{T}^n \times \mathbb{R}^n$, Mémoire n^0 48, Supplément au bulletin de la S.M.F., Tome 120, fasc. 1

Arnold, V. (1963) Small denominators and problems of stability of motion in classical and celestial mechanics, Russ. Math. Surv., 18

Arnold, V. (1965) Sur une propriété topologique des applications globalement canoniques de la méchanique classique, C.R. Acad. Sci. Paris Série I Math. 261, Groupe 1

Arnold, V. (1978) Mathematical Methods of Classical Mechanics, Springer-Verlag

Arnold, V. (1983) Geometric methods in the theory of ordinary differential equations, Springer–Verlag

Bangert, V. (1988) Mather sets for twist maps and geodesics on tori, Dynamics reported 1

Bangert, V. (1989) Minimal Geodesics, Ergod. Th. & Dynam. Sys., 10

Bernstein, D. and Katok, A. (1987) Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonians, Invent. Math. 88

Birkhoff (1920) Surface transformations and their dynamical applications. Collected Math. Papers, vol. 2

Bialy, M. and Polterovitch, L. (1992) Hamiltonian systems, Lagrangian tori and Birkhoff's Theorem, Math. Ann. 292

Boyland, P. (I996) New Dynamical Invariants on Hyperbolic Manifolds, Preprint

Bost, J–B. (1986) Tores invariants des systèmes dynamiques Hamiltoniens, Séminaire Bourbaki no. 639, Astérisque 133-134

Boyland, P. (1988) Rotation sets and Morse decomposition for twist maps, Ergod. Th. & Dynam. Sys. 8

Boyland, P. and Golé, C. (1996a) Dynamical stability in Lagrangian systems, in Hamiltonian Systems with Three or More Degrees of Freedom, C. Simó, editor, Kluwer Acad. Publ., Dordrecht, Holland.

Boyland, P. and Golé, C. (1996b) Lagrangian systems on hyperbolic manifolds, to appear in Ergod. Th. & Dynam. Sys.
(See also www.math.sunysb.edu/cgi-bin/preprint.pl?ims96-1a)

Candel, A. & de la Llave, R. (1997) On the Aubry-Mather theory in statistical mechanics, Comm. Math. Phys. 192 no. 3

Guillou, L. (1992) A generalized translation theorem for free homeomorphisms of surfaces, Preprint No. 217 of Institut Fourier, Grenoble
Hall, G. (1984) A topological version of a theorem of Mather on twist maps, Ergod. Th. and Dynam. Sys. 4
Hall, G. (1989) A topological version of a theorem of Mather on shadowing in monotone twist maps. Dynamical systems and ergodic theory (Warsaw, 1986), Banach Center Publ., 23, PWN, Warsaw
Hedlund (1932) Geodesics on a two-dimensional Riemannian manifold with periodic coefficients, Ann. Math. 33
Herman, M. (1983) Sur les courbes invariantes par les difféomorphismes de l'anneau, vol. I, Asterisque, 103-104
Hermann, M. (1990) Inegalités à priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques, Publ. Math. I.H.E.S. 70
Herman, M. (1992a) On the dynamics of Lagrangian tori invariant by symplectic diffeomorphisms. Progress in variational methods in Hamiltonian systems and elliptic equations (L'Aquila, 1990), Pitman Res. Notes Math. Ser., 243, Longman Sci. Tech., Harlow
Herman, M. (1992b) Dynamics connected with indefinite normal torsion, in Twist mappings and their applications, IMA volume 44, Springer-Verlag
Hirsch, M. (1988) Stability and convergence in strongly monotone dynamical systems J. reine angew. Math. 383
Hirsch, M. and Smale, S. (1974) Differential equations, dynamical systems and linear algebra, Academic Press
Katok, A. (1982) Some remarks on Birkhoff and Mather twist map theorems, Ergod. Th. & Dynam. Sys., 2
Katok, A. (1992) Minimal orbits for small perturbations of completely integrable Hamiltonian systems, in Twist mappings and their applications, IMA volume 44, Springer-Verlag
Katznelson, Y. and Ornstein, D. (???) Aubry-Mather via trimming ???
Kolmogorov, A. (1954) The general theory of dynamical systems and classical mechanics, address to the 1954 International Congress of Mathematicians, English translation in the appendix of Abraham & Marsden (1985)
Kook, H. and Meiss, J. (1989) Periodic orbits for Reversible, Symplectic Mappings, Physica D 35
Hasselblat, B. and Katok, A. (1995) Hasselblat & Katok (1995) (1995) Introduction to the modern theory of dynamical systems, Cambridge University Press.
Klingenberg, W. (1982) Differential geometry, Kleuwer ???
Koch, H., de la Llave, R. & Radin, C. (1994) Aubry-Mather theory for functions on lattices. Discrete Contin. Dynam. Systems 3, no. 1
Delshams, A., de la Llave, R., Seara, T. M. (1998) A geometric approach to the existence of orbits with unbounded energy in generic periodic perturbations by a potential of generic geodesic flows of $\mathbb{T}^2$. Preprint 98-591 available at http://rene.ma.utexas.edu:80/mp_arc/.
de la Llave, R. (1999) Variational methods for quasi–periodic solutions of partial differential equations, preprint
Lang, S. (1983) Real Analysis, Addison-Wesley
Laudenbach & Sikorav (1985) Persistance d'intersection avec la section nulle au cours d'une isotopie hamiltonienne dans un fibré cotangent, Invent. Math. 82
LeCalvez, P. (1990) IMPA ???
LeCalvez, P. (1991) Propriétés dynamiques des difféomorphismes de l'anneau et du tore, Astérisque 204
LeCalvez, P. (??? GCs) Find his paper where GCs are constructed ???.
Levi M. (???) His Aubry-Mather theorem on the Hedlund example (Erg. Th. Dyn. Sys (1997)???)
Lochak, P. (1992) Canonical perturbation theory via simultaneous approximation, Russ. Math. Surv. 47:6
McDuff, D. and Salamon, D. (1996) Introduction to Symplectic Topology, Oxford Mathematical Monographs
MacKay, R. (???) Twist maps, renormalization ???
MacKay, R. and Meiss, J. (1983) Linear stability of periodic orbits in Lagrangian systems, Phys. Lett. **98A**, 92
MacKay, R., Meiss, J. & Percival, I. (1984) Transport in Hamiltonian systems, Physica 13D
MacKay, R., Meiss, J. & Percival, I. (1986) Resonances in area–preserving maps, Physica 27D
MacKay, R., Meiss, J. and Stark, J. (1989) Converse KAM theory for symplectic twist maps, Nonlinearity 2
MacKay, R. & Muldoon, M. GCs...

Mañe, R. (1987) Ergodic theory and differentiable dynamics, Springer-Verlag
Manẽ, R. (1991) Global variational methods in conservative dynamics, 18^o Colóquio Brasileiro de Matemática, IMPA
Mather, J. (1982) Existence of quasiperiodic orbits for twist homeomorphisms, Topology 21
Mather, J. (1984) Non-existence of invariant circles, Ergod. Th. & Dynam. Sys. 2
Mather, J. (1986) A criterion for the non-existence of invariant circles Publ. Math. I.H.E.S. 63
Mather, J. (1991b) Action minimizing invariant measures for positive definite Lagrangian systems, Math. Z. 207
Mather, J. (1993) Variational constructions of connecting orbits, Ann. Inst. Fourier, Grenoble 43, No. 5
Mather, J. (diffusion???) Recent work on diffusion ???
Meiss, J. (1992) Symplectic maps, variational principles, and transport, Rev. Mod. Phys. 6, No. 3
Milnor, J. (1969) Morse theory, Princeton University Press
Moser, J. (1962) On invariant curves of area–preserving mappings of an annulus, Nachr. Akad. Wiss. Gŏttingen, Math. Phys. K1.II (see also the next reference)
Moser, J. (1973) Stable and random motions in dynamical systems, Princeton University Press
Moser, J. (1977) Proof of a generalized form of a fixed point theorem due to G.D. Birkhoff, Lecture Notes in Mathematics, Vol. 597: Geometry and Topology, pp. 464-494, Springer
Moser, J. (1986) Monotone twist mappings and the calculus of variations, Ergod. Th. and Dyn. Sys., Vol 6
Nekhoroshev, M. (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems, Russ. Math. Surv. 32, No. 6
Percival (1979) A variational principle for invariant tori of fixed frequency, J. Phys. A 12
Pöshel, J. (1993) Nekhoroshev estimates for quasi-convex Hamiltonian systems. Math. Z. 213, No. 2
Poincaré, H. (1885) Sur les courbes définie par les équations différentielles, Chapter XV, Journal de Mathématiques pures et appliquées, 4^e série, t.1 (this and the following two references can be also found in the "Œuvres complètes de Henri Poincaré", Gauthier Villar, (1951))
Poincaré, H. (1890) Sur le problème des trois corps et les équations de la dynamique, Acta Math. 13
Poincaré, H. (1895) Analysis situs, J. de l'Écolde Polyt. (2) 1
Poincaré, H. (1912) Sur un théorème de géométrie, Rand. Circ. Math. Palermo, 33
Poincaré, H. (1892) Les méthodes nouvelles de la méchanique céleste (3 volumes), Gauthier–Villars, Paris
Robinson, C. (1970) Generic properties of conservative systems, Am. Jour. Math., Vol XCII, No. 3
Robinson, C. (1994) Dynamical Systems, CRC Press
Rom-Kedar, V. & Wiggins, S. (1990) Transport in two-dimensional maps, Arch. Rat. Mech. & Anal. Vol 109, Number 3
Rüssmann, H. (1970) Kleine Nenner I: Über invarianten Kurven differenzierbarer Abbidulgen eines Kreisringes, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.
Salamon, D. (???) Conley index....????
Salamon, D. (1990) Morse Theory, the Conley index and Floer homology, Bull. London Math. Soc. 22
Siburg, K.F. (1995) Symplectic Capacities and Bi-invariant Metrics on Hamiltonian Diffeomorphisms, Thesis Diss. Nr.11110, ETH Zürich
Sikorav (1986) Sur les immersions lagrangiennes dans un fibré cotangent admettant une phase génératrice globale, C.R. Acad. Sci. Paris, t. 302, Série I. n^0 3
Slijepčević, S. (1999a) Monotone gradient dynamics and Mather's shadowing, Nonlinearity 12, no. 4
Slijepčević, S. (1999b) Accelerating orbits of twist diffeomorphisms on a torus, Glas. Mat. Ser. III 34(54) (1999), no. 1
Spivak, M. (1970) A comprehensive introduction to differential geometry, Vol 1, Publish or Perish
Tabacman, E. (1993) Ph.D. Thesis, University of Minnesota.
Tabacman, E. (1995) Variational computation of homoclinic orbits for twist maps, Phys. D 85 (1995), no. 4
Tangerman, F, and Veerman, JP. (1989) Renormalization of Aubry-Mather Cantor sets, J. Statist. Phys. 56, no. 1-2
Tangerman, F, and Veerman, JP. (1990a) On Aubry-Mather sets. Phys. D 46, no. 2
Tangerman, F, and Veerman, JP. (1990b) Asymptotic geometry of hyperbolic well-ordered Cantor sets, J. Statist. Phys. 59 (1990), no. 1-2
Tangerman, F. and Veerman, JP. (1991) Intersection properties of invariant manifolds in certain twist maps, Commun. Math. Phys. 139
Theret, D. (1999) A complete proof of Viterbo's uniqueness theorem on generating functions, Topology Appl. 96 , no. 3

Weinstein, A. (1979) Lectures on Symplectic Manifolds, CBMS series no. 29, A.M.S.
Viterbo, C. (1987) Intersection de sous variétés Lagrangiennes, fonctionnelles d'action et indice de systèmes Hamiltoniens, Bull. Soc. math. France 115
Viterbo, C. (1992) Symplectic topology as the geometry of generating functions, Math.Ann. 292
Weinstein, A. (1979) Lectures on symplectic manifolds, C.B.M.S. Conf. Series, no.29, AMS
Wiggins, S. (1990) On the geometry of transport in phase space. !. Transport in k-degree-of -freedom Hamiltonian systems, $2 \leq k < \infty$., Physica D 44
Yoccoz, J.-C. (1992) Travaux de Herman sur les tores invariants, Séminaire Bourbaki no. 754, Astrisque No. 206

What about: Delshams, A.(E-UPB-A1); Ramírez-Ros, R.(E-UPB-A1) Exponentially small splitting of separatrices for perturbed integrable standard-like maps. (English. English summary) J. Nonlinear Sci. 8 (1998), no. 3, 317–352. (the first rigorous proof of exponentially small splitting? write Amadeu...)

Content

1/2/00

Index

Index

1/2/00