INTRODUCTION

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 perturbs 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 winding around the cylinder. 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).

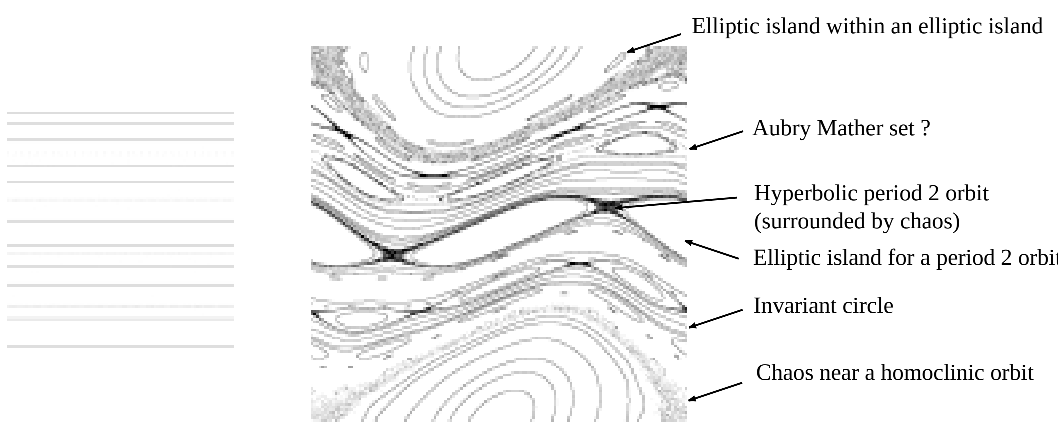

Fig. 1.1. The different dynamics in the standard map: the left hand side shows a selection of orbits for the completely integrable F_0, all on invariant circles. The right hand side displays orbits for F_ϵ with $\epsilon = .817$.

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 (1984) says that no invariant circle survives. However, the deep theory of Kolmogorov-Arnold-Moser (KAM, see Chapter 6) 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 2), 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.

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 are indistinguishable in dimension 2). The vertical fibers $\{x = x_c\}$ are still sheared, in a way made precise in Chapter 4. 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 $\{y = y_c\}$, and induces a translation by the vector y_c on each torus. 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 for $small$ symplectic perturbations f_ϵ of f_0. 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 5). What about orbits of irrational rotation vector? There are counter-examples to a $full$ analog of the Aubry-Mather theorem in higher dimensions, in which the rotation vectors of $action\ minimizing\ orbits$ can be sharply restricted. 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 9.

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 $\int_\gamma L dt$. 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 7. 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 6). 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.

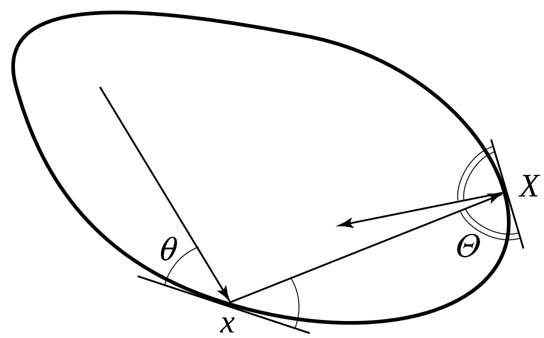

Fig. 2.1. In a convex billiard, the point x and angle θ at a rebound uniquely and continuously determines the next point X and incidence angle Θ.

The map f can be seen as a limit of section maps for the geodesic flows[1] of a sphere that is being flattened until front and back are indistinguishable. The boundary of the billiard is the fold of the flattened sphere (not so round in our illustration). Now, 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 of crossing makes correspond the next crossing point and angle. Seen in

[1] 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.

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).

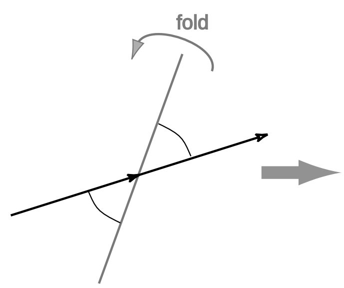

Fig. 2.2. The law of reflexion as a consequence of folding.

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 6). If we inflate our billiard back a little, these polygonal lines become *broken geodesics* on the partially inflated sphere. 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 some non zero angle, generally. 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. 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 closed 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 7, and again in Chapter 10.

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)).

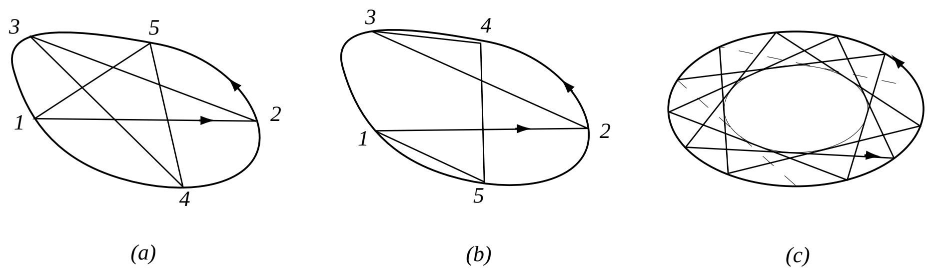

Fig. 2.3. Different polygonal configurations in billiards: (a) is of period 5, rotation number 3/5 and is cyclically ordered. (b) is also of period 5, but of rotation 1/5 and is not cyclically ordered. Note that neither (a) nor (b) represent orbits since the law of reflexion is not satisfied. (c) is a configuration corresponding to an orbit on an invariant circle for the completely integrable elliptic billiard map. Its rotation number is presumably irrational.

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 key to both proofs of the Aubry-Mather theorem we give in this book. In the second proof, this order property imparts some monotonicity on the gradient flow of the action. Unfortunately, there is no natural order for orbits of higher dimensional twist maps. But the same kind of ordering exists in higher dimensional non dynamical models that generalize the Frenkel-Kontorova setting (see Chapter 9).

3. An Ancestor of Symplectic Topology

At the end of his life, Poincaré 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 (1912) gave an incomplete proof of this theorem. In a moving introduction, he states that he had never done that before, and that it would have been wiser for him to let rest this important problem on which he had spent almost two years of work, to come back and finish it later. But, as he points out: “à mon age, je ne puis y répondre[2]”, and indeed, he died in year. Birkhoff (1913) gave a substantially different proof, which was also somewhat incomplete as to the existence of at least *two* fixed points[3]. Since then, a number of new proofs have appeared (Brown & Von Neuman (1977), 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 (1991) we present in Section 7 and the modern theory of symplectic topology.

Sketch of Proof of the Poincaré-BirkhoffTheorem. 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

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

[2] at my age, I cannot count on it

[3] it did prove the existence of at least one: he had overlooked the possibility of fixed points of index 0. Birkhoff (1925) contains a proof of a more topological version of the theorem, in which he corrected the problem of his first proof. Some mathematicians were still unsure about the validity of his proof. Brown & Von Neuman (1977) gives a rigorous version of his proof.

which is the set of points that only move up or down under the map[4]. 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 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. $\square$

Generating Functions. 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. 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 closed symplectic manifolds by a similar argument.

Intersections of Lagrangian Manifolds. The above scheme of proof can be rephrased in terms of intersections of Lagrangian manifolds. 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. Both these sets are prototypical Lagrangian manifolds (see Appendix 2). The function w is called a generating (phase) function for the manifold $F(\Gamma)$. Hence

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

the proof of Poincaré's geometric theorem is reduced, in this simple case, to the proof of intersection of two Lagrangian manifolds. Important theorems ($eg.$ the Arnold Conjecture) 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. Two problems arise in general: 1) to find a generating function for a Lagrangian manifold which is not a graph and 2) to estimate the number of critical points of this generating function. In this book, we approach the first problem by the method of decomposition of symplectic maps in twist maps (in the proof of Poincaré's theorem in Chapter 1, and its generalization to higher dimension, Theorem 43.1), a method very much related to that of "broken geodesics" (see Chapter 10). As for the second problem, we use Conley's theory here, and its refinements by Floer in his work on the Arnold's Conjecture.