14

Flows on surfaces and related dynamical systems

In this chapter we study a class of continuous-time dynamical systems with very low-dimensional behavior according to the description given in Chapter 10, namely, smooth flows on closed compact surfaces. We will also pay attention to flows on surfaces with boundary such as the closed disc or the cylinder and on open surfaces such as the plane. This, in particular, will allow us to treat semilocal problems. Another natural object associated with such flows are Poincaré maps induced on transversals to the flow. If the flow preserves a nonatomic measure positive on open sets (for example, an area) then such Poincaré return maps are topologically conjugate to a locally isometric map with finitely many discontinuities. The term "interval exchange transformation" gives a visual description of such a map.

In general, the asymptotic behavior of flows on surfaces is characterized by slow orbit growth but they have less uniform types of recurrence and of statistical behavior than invertible one-dimensional maps studied in Chapters 11 and 12. The former aspect is closely related to the fact that both orbits and one-dimensional transversals to a flow locally separate the surface; the latter is due primarily to the more complicated topology of surfaces of genus greater than one compared to the circle (and the torus) and to a lesser extent to the effects of time change. Typical manifestations of this type of complexity, intermediate between the simple behavior of our first group of examples (Sections 1.3–1.6) and the circle diffeomorphisms on the one hand, and that of the examples with positive topological entropy (Sections 1.7–1.9, 5.4, 9.6) on the other, are finiteness results for the number of nontrivial orbit closures (Theorem 14.6.3) and nonatomic ergodic invariant measures (Theorem 14.7.6) for flows on surfaces of genus greater than one, which replace uniqueness of a minimal set (Proposition 11.2.5) and unique ergodicity (Theorem 11.2.9) for circle homeomorphisms.

Even though most of these results hold in full generality we prove them under simplifying assumptions, such as finiteness and not too degenerate structure of the fixed points and preservation of area or a measure positive on open sets. This is justified by the facts that first unlike in the case of circle maps these assumptions do not restrict us to essentially trivial situations and indeed convey the entire possible complexity of flows on surfaces, and second that this case appears in several interesting situations such as billiards in polygons with rational angles.

1. Poincaré–Bendixson theory

a. The Poincaré–Bendixson Theorem. We begin with the study of flows on those surfaces whose topology allows only simple types of recurrent orbits, that is, fixed points and periodic orbits. This does not mean that the global orbit structure of any flow on such a surface is trivial, but only that the complexity of that structure would have be due to the combinatorial picture of fixed points, periodic orbits, and saddle connections rather than to recurrent behavior. The sphere, the plane, and the disc are prime examples of surfaces with that property, but this class also includes the cylinder, the Möbius strip, and the projective plane. The arguments are based on the Jordan Curve Theorem A.5.2. We obtain the following:

Theorem 14.1.1. (Poincaré–Bendixson Theorem) *Let M be a surface that is an open subset of the sphere S^2 or the projective plane. Let X be a C^1 vector field on M. Then all positively or negatively recurrent orbits are periodic. Furthermore, if the ω-limit set of a point contains no fixed points, then it consists of a single periodic orbit.*

Proof. Suppose first that M is a subset of the sphere. Denote by φ^t the flow generated by X and suppose p is positively recurrent and nonperiodic. Take a short transversal γ at p and let t be the smallest positive number for which $\varphi^t(p) \in \gamma$. Then the union of the orbit segment $\{\varphi^s(p)\}_{0\leq s\leq t}$ and the piece of γ between p and $\varphi^t(p)$ is a simple closed curve $\mathcal{C}$ called a *pretransversal* (because we shall later use such curves to construct transversals). By the Jordan Curve Theorem A.5.2 the complement of $\mathcal{C}$ consists of two disjoint open sets A and B. We may label them such that near γ the flow goes from A to B. This implies that the positive semiorbit of $\varphi^t(p)$, hence the ω-limit set $\omega(p)$ of p, is in B. Since p is recurrent we have $A \ni \varphi^{-\epsilon}(p) \in \mathcal{O}(p) \subset \omega(p) \subset B$, a contradiction.

If M is a subset of the projective plane then there is an orientable double cover of M. If $p \in M$ is a positively recurrent nonperiodic point, then consider the two points p_1 and p_2 that cover it. The orbit of p_1 under the flow generated by the lift of the vector field X accumulates on $\{p_1, p_2\}$. If it accumulates on p_1 then we are done by the previous argument. Otherwise it accumulates on p_2 and we can construct a pretransversal near p_2, which again leads to a contradiction.

Now consider the ω-limit set W of a point p and assume that it contains no fixed points. By Corollary 3.3.7 there are recurrent points in W. By the above these are periodic. Thus let $q \in W$ be a periodic point. Note that in the case of the projective plane the lift of q to the orientable double cover is still periodic, so we may assume that M is orientable. Consider a small transverse segment

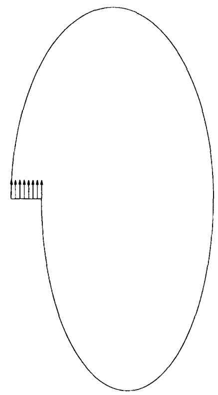

FIGURE 14.1.1. A pretransversal

γ containing q. By continuity the return map to this segment is defined on a neighborhood of q in γ. Take a one-sided neighborhood I of q small enough so that the first point $\varphi^t(p)$ in γ is not in I, but infinitely many of these returns are. Parameterizing this neighborhood by $[0, \delta)$ gives a continuous map f from an interval $[0, \delta)$ to an interval $[0, \delta')$ that fixes 0. The orbit of p provides infinitely many $x \in (0, \delta)$ for which $f(x) < x$, so either $f(x) < x$ for all $x \in [0, \delta)$ or $[0, \delta)$ contains a fixed point y. The latter case is impossible, since the interval $[0, y]$ would be invariant under f and hence there would be an invariant annulus for the flow that separates the orbit of q from that of p, so $q \notin \omega(p)$. But if $f(x) < x$ then all $x \in (0, \delta)$ are positively and monotonically asymptotic to 0. Since the return times to I are bounded this means that the orbit segments of p between successive returns converge to the orbit of q, so $\omega(p)$ coincides with the orbit of q. □

b. Existence of transversals.

Definition 14.1.2. A *transversal* to a vector field on a surface is a simple closed curve such that the vector field is nowhere tangent to the curve.

Let τ be a transversal to a vector field X and fix an orientation of τ. Then at each point of τ we can define the angle between τ and X. This angle is either in $(0, \pi)$ or in $(0, -\pi)$ for all points.

Proposition 14.1.3. *Let M be a surface with a Riemannian metric and X a C^1 vector field on M with a positively recurrent nonperiodic orbit. Then for every $\epsilon > 0$ there exists an oriented transversal τ such that the angle between X and τ is in $(0, \epsilon)$ at every point.*

We note that the proof is easier in the case of orientable surfaces, but that nonorientability is only a minor complication.

Proof. We begin with a construction which is a more careful version of our construction of pretransversals. Naturally, since we are on an arbitrary surface we cannot use the Jordan Curve Theorem A.5.2 and we do not get a contradiction as in the proof of Theorem 14.1.1. Let p be a positively recurrent point and consider a *flow box*, that is, a neighborhood $\mathcal{U}$ of p on which there are C^1 coordinates (x, y), $-\epsilon \le y \le \epsilon$, in which $X = \frac{\partial}{\partial y}$ and $p = 0$. The boundary curve given by $y = -\epsilon$ is called the *base* of the flow box, and the curve $y = \epsilon$ the *roof*.

Denote the flow generated by X by φ^t. Since p is positively recurrent, but not periodic, there are infinitely many values of t for which $\varphi^t(p)$ is on the line $y = 0$ in $\mathcal{U}$. Since X does not vanish along the orbit of p, we can choose a vector field Y along the positive semiorbit of p such that (X, Y) is a frame, that is, X and Y are linearly independent at each point. This defines an orientation for all points of the orbit of p. For any two points of $\varphi^t(p)$ in $\mathcal{U}$ we can compare these orientations. Given $\delta > 0$ let us call a time t_0 a *closest-return* time if the point $\varphi^{t_0}(p)$ is on the line $y = 0, |x| < \delta$ in $\mathcal{U}$ and closer to 0 than any point $\varphi^t(p)$ on the same line for any $t \in (0, t_0)$. We would like to have infinitely many closest returns whose orientation agrees with that at p. If that is not the case then the orientations at p and $\varphi^t(p)$ differ for infinitely many closest returns, so we can consider two successive closest-return times t_0 and t_1 such that the orientations at $\varphi^{t_0}(p)$ and $\varphi^{t_1}(p)$ agree and the distance between $\varphi^{t_0}(p)$ and $\varphi^{t_1}(p)$ is as small as we please. Replacing p by $\varphi^{t_0}(p)$ then puts us in the first case. Thus we may assume that the point p has a closest return $\varphi^{t_0}(p)$ at which the orientation coincides with that at p. Notice that the preceding argument is only needed to take care of the case of nonorientable surfaces.

Consider now a narrow strip around the orbit segment of p for $t \in [0, t_0]$. This strip may be assumed to be orientable since the frame (X, Y) along the orbit of p can be extended to a frame on a small neighborhood of the orbit segment. Using the Riemannian structure we can thus define a rotated vector field $Z = R_\theta X$ as the vector field that has angle θ with X. If the angle is small enough and has the right sign then the orbit of p under the flow ψ_θ generated by Z stays within the strip and returns to a point $\psi_\theta^{t'}(p)$ between p and $\varphi^{t_0}(p)$ on the line $y = 0$. Consider the curve c defined by connecting the points $\psi_\theta^{t_1}$ and $\psi_\theta^{t_2}(p)$ with a straight line in $\mathcal{U}$, where t_1 and t_2 are chosen such that $\psi_\theta^t(p) \in \mathcal{U}$ for $t \in [0, t_1]$ and $t \in [t_2, t']$. If t_1 is not too small and t_2 is not too close, then for sufficiently small δ the angle between $\dot{c}$ and X is less than ϵ where defined. Thus we can take a smooth curve τ sufficiently close to c such that the angle between $\dot{\tau}$ and X is between 0 and ϵ along τ. This is the desired transversal. □

We would like to see where the return map to such a transversal is defined. This construction clearly yields a transversal τ such that there is at least one point $q \in \tau$ that returns to τ in positive time. Thus the return map to τ is well defined and continuous on a nonempty open subset of τ. The following result gives a way to show that this is a large set in some cases. It applies not only to closed transversals, but to transverse segments as well. Notice that the set of points on the transversal that return to it is open and hence a union of disjoint intervals.

Proposition 14.1.4. *Let M be a closed surface with a C^1 vector field X. Suppose τ is a transversal, not necessarily closed. If τ_0 is the set of points returning to τ and $p \in \partial\tau_0$ an endpoint of an interval in τ_0 that is not an endpoint of τ and does not return to an endpoint of τ, then the ω-limit set $\omega(p)$ consists of fixed points of X.*

Proof. Denote by φ^t the flow generated by X and let us suppose that there is a point $y \in \omega(p) \smallsetminus \operatorname{Fix}(\varphi^t)$. Observe that in this case the orbit of p has infinite length since there is an $\epsilon > 0$ and a neighborhood of y (for example, a flow box) to which the orbit returns infinitely many times and such that upon each return the orbit segment in the neighborhood has length ϵ. Since M is compact, the orbit of p thus has a nonperiodic recurrent limit point. This means that for $\epsilon > 0$ there are times $0 < t_1 < t_2$ such that $d(\varphi^{t_1}(p), \varphi^{t_2}(p)) < \epsilon$ and the orientations at $\varphi^{t_1}(p)$ and $\varphi^{t_2}(p)$ (as defined in the previous proof) coincide. By closing $\{\varphi^t(p) \mid t_1 \le t \le t_2\}$ by a short transverse curve we obtain a pretransversal. So from the orbit of p we have just produced pretransversals with arbitrarily short transverse pieces.

Note that given two segments τ_1 and τ_2 transverse to X for which there is an interval $I \subset \tau_1$ of points whose positive orbits intersect τ_2, there is a flow box consisting of orbit segments beginning on I and ending on τ_2. It is useful to note that any closed transversal that intersects an orbit segment of a flow box must intersect the base or roof or else traverse the entire flow box, that is, intersect every orbit segment in the flow box. The same holds for the transverse segment of a pretransversal as long as its ends are not in the flow box.

Given $q \in I$ let r be the midpoint (in I) between p and q and denote by J the interval with endpoints q and r. Let d be the minimum width of the flow box with base J and roof in τ (that is, the infimum of lengths of curves intersecting all orbit segments in the flow box). As we have observed, we can construct a pretransversal from the orbit of p whose transverse part γ has length less than $d/2$ and is disjoint from τ. Note that the positive semiorbit of every point p' of I sufficiently close to p must intersect γ (either at time approximately t_1 or at time approximately t_2). The interval with endpoints p and p' is the base of a flow box consisting of orbit segments ending on γ. Except for the base, τ is disjoint from this flow box, since otherwise τ would intersect all orbit segments in the flow box, in particular that of p. But by assumption p does not return to τ.

Consider now the flow box whose base is the interval in τ with endpoints q and p' and whose roof is in τ. We may assume that this interval contains the midpoint r between p and q. The previous argument shows that γ intersects this flow box and hence traverses it completely (since it is disjoint from τ).

Hence the length of γ exceeds d by choice of d. On the other hand we chose γ to have length less than $d/2$. This contradiction ends the proof. □

Corollary 14.1.5. *If X is a fixed-point-free vector field on a closed surface and τ a closed transversal, then the return map to τ is either defined on all of τ or not defined at all.*

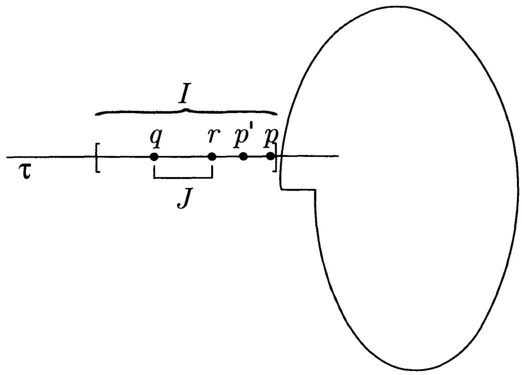

FIGURE 14.1.2. Arrangement of the transversals

Proposition 14.1.6. *Let X be an area-preserving vector field on a closed compact surface M all of whose fixed points are centers and (possibly multiple) saddles. Let τ be a not necessarily closed transversal to X. Then the return map on τ is defined and continuous at all but finitely many points and at those points both one-sided limits exist.*

Proof. First, the set of fixed points of X is finite since saddles are isolated and M is compact. There are finitely many positive semiorbits whose ω-limit sets consist of fixed points. Those are incoming (stable) separatrices of the saddles. In fact, the ω-limit set of any separatrix consists of a single saddle.

By Proposition 4.1.18(1) applied to a time-1 map the set of recurrent points is dense in M; hence its intersection with any flow box $\mathcal{U}$ around $p \in \tau$ is dense in $\mathcal{U}$. Since the set of recurrent points is flow-invariant, the set τ_0 of points in τ that return to τ is dense in τ. Recall that τ_0 is open and hence consists of intervals. Hence by Proposition 14.1.4 the positive semiorbit of any endpoint of any component of τ_0 must be an incoming separatrix of one of the fixed points. Thus τ_0 consists of finitely many intervals, and since it is dense in τ, its complement is finite.

Note that on each interval of τ_0 the return map is injective, hence monotone, so the one-sided limits exist at the endpoints. □

Remark. The only way invariance of the area was used was through Proposition 4.1.18, which can be applied if the flow has an invariant measure whose support is the whole surface, that is, which is positive on open sets. However, in the area-preserving case extra information can be obtained about the return map.

Corollary 14.1.7. *Under the assumptions of the previous proposition there exists a smooth parameter on τ such that the return map to τ is an isometry on each component of τ_0.*

Proof. Consider the smooth invariant measure on τ induced by the (invariant) area (cf. Proposition 5.1.11). Pick an initial point $p \in \tau$ and an orientation on τ. Then parameterize points $x \in \tau$ by the (signed) measure of the interval between p and x. □

In Section 14.5 we will study the return maps that appear in this corollary.

Exercises

14.1.1*. *Show that there is a C^r-open dense set of C^r vector fields on a compact closed surface that have only finitely many fixed and periodic orbits, all of them hyperbolic.*

14.1.2. *Show that for any area-preserving vector field on the sphere there is a dense invariant set that consists of fixed points and periodic orbits.*

14.1.3. *Show that the Poincaré–Bendixson Theorem 14.1.1 holds for any flow generated by a C^0 vector field so long as the latter is uniquely integrable.*

2. Fixed-point-free flows on the torus

a. Global transversals. According to the Poincaré–Hopf Index Theorem 8.6.6 the only compact surfaces (without boundary) that admit fixed-point-free flows are the torus and the Klein bottle. It turns out that on the Klein bottle nontrivial recurrence is impossible (see Exercise 14.2.3) and that on the torus any flow with nontrivial recurrence is equivalent to a flow under a function built from a circle diffeomorphism.

We will use the following observation. Consider any compact orientable surface M with a Riemannian metric. Then given a vector field X on M there is a one-parameter family of vector fields $R_\theta X$ obtained by rotating $X(p)$ by an angle θ at every point p. By orientability this is well defined. Obviously the set of fixed points of $R_\theta X$ coincides with that of X.

Proposition 14.2.1. *A fixed-point-free C^1 flow on the torus $\mathbb{T}^2$ admits a closed transversal.*

Proof. Fix a Riemannian metric on $\mathbb{T}^2$ and consider the vector field $Z = R_{\pi/2}X$ perpendicular to X. If Z has a periodic orbit then this gives a closed transversal to X. Otherwise Z has a positively recurrent nonperiodic orbit and we use Proposition 14.1.3 to obtain an oriented transversal to Z that makes an angle less than $\pi/4$ with Z. But this then is also a transversal to X since the angle with X cannot be zero. □

Proposition 14.2.2. *If a fixed-point-free flow with a nonperiodic recurrent orbit on a closed surface admits a closed transversal then every orbit intersects the transversal.*

Proof. We will show that the set of points whose orbit intersects a transversal τ is open and closed in $\mathbb{T}^2$. Note first that by Corollary 14.1.5 the return map is defined on the entire transversal. Note that by compactness of τ the return times are bounded. Thus the set of points whose orbit intersects τ is a finite union of images of the closed set $\bigcup_{t\in[-\epsilon,\epsilon]}\varphi^t(\tau)$, hence is closed. Likewise, however, this set is also a finite union of open sets $\bigcup_{t\in(-\epsilon,\epsilon)}\varphi^t(\tau)$, hence is open. □

These two results now yield the advertised equivalence

Corollary 14.2.3. *A fixed-point-free C^1 flow on $\mathbb{T}^2$ with a nonperiodic recurrent orbit is smoothly conjugate to the flow under a function over an orientation-preserving circle diffeomorphism.*

Proof. By Proposition 14.2.1 the flow admits a transversal τ and by Proposition 14.2.2 every orbit intersects τ. If we parameterize τ by an angle $\theta \in S^1$, then the return map to τ defines a circle diffeomorphism f. If the return time of θ is $h(\theta)$, then we can coordinatize $\mathbb{T}^2$ by (θ, y) with $0 \leq y < h(\theta)$ and in these coordinates the vector field is $\frac{\partial}{\partial y}$, that is, the flow is a special flow. □

This close relation to circle maps will allow us to apply results from Chapters 11 and 12. A first instance is the following classification.

Proposition 14.2.4. *A C^2 fixed-point-free flow on $\mathbb{T}^2$ with a nonperiodic recurrent orbit is topologically conjugate to a time change of a linear flow.*

Proof. By Proposition 14.2.1 there is a transversal, which we may assume to be smooth, so the flow is conjugate to a special flow over a C^2 circle diffeomorphism. By Theorem 12.1.1 this diffeomorphism is topologically conjugate to a rotation, so via Corollary 14.2.3 the flow is topologically conjugate to a special flow over a rotation, that is, a time change of a linear flow on $\mathbb{T}^2$. □

The complications we found in the theory of circle maps of regularity less than C^2, namely, the Denjoy example (cf. Proposition 12.2.1), arise in this situation as well. Namely, the special flow over a Denjoy map from Section 12.2 is a flow on the torus which exhibits Denjoy-type behavior and in particular is not topologically conjugate to a special flow over a rotation.

b. Area-preserving flows. In certain cases the topological conjugacy obtained in Proposition 14.2.4 is actually a smooth conjugacy. We first consider area-preserving flows, where the Poincaré Recurrence Theorem 4.1.19 provides the recurrence needed in the last two statements. .

Proposition 14.2.5. *Let φ^t be a C^k flow of $\mathbb{T}^2$ preserving a C^r area element and let $n = \min(k, r+1)$. Then φ^t is C^n conjugate to a special flow over a circle rotation.*

Proof. Note first that the C^∞ transversal τ obtained in Proposition 14.2.1 has a $C^{\min(k,r)}$ length parameter invariant under the return map by Proposition 5.1.11. Thus the return map is C^n conjugate to a circle rotation by Proposition 12.4.4. By extending this conjugacy as in the proof of Corollary 14.2.3 we obtain a C^n conjugacy to a special flow. □

Note that a given linear flow can be represented as a suspension of different rotations by choosing different transversals. It is easy to see, however, that the rotation number depends only on the homotopy type of the transversal. In particular consider the flow T_t^ω and the transversal $\tau_{k,l}$ that lifts to the straight line with slope l/k. Then consider a linear transformation given by $A = \begin{pmatrix} k & m \\ l & n \end{pmatrix}$ with integers m, n such that $\det A = 1$. The rotation number with respect to the transversal $\tau_{k,l}$ is the rotation number obtained from $T_t^{A^{-1}\omega}$ using the horizontal transversal $\tau_{1,0}$, that is, the reciprocal of the slope of $A^{-1}\omega$, which is given by $\frac{n\omega_1 - m\omega_2}{-l\omega_1 + k\omega_2}$. Notice that (m, n) is unique up to integer multiples of (k, l), so this expression for the rotation number is well defined modulo 1. In particular the rotation numbers ρ and ρ' obtained from two different transversals are related by a linear fractional transformation

$$\rho' = \frac{a\rho + b}{c\rho + d},$$

where $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an integer matrix with unit determinant.

Recall (Definition 2.8.1) that an irrational number α is called *Diophantine* if there exist $k, r > 0$ such that for any nonzero $p, q \in \mathbb{Z}$ we have $|q\alpha - p| > kq^{-r}$. This property is invariant under linear fractional transformations.

Lemma 14.2.6. *If α is a Diophantine irrational and $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ a nonsingular integer matrix then $\dfrac{a\alpha + b}{c\alpha + d}$ is Diophantine as well.*

Proof. Note that by assumption a and c cannot both vanish. Thus we find

$$\left|q\Big(\frac{a\alpha + b}{c\alpha + d}\Big) - p\right| = \left|\frac{(qa - pc)\alpha + (qb - pd)}{c\alpha + d}\right| > \frac{k|qa - pc|^{-r}}{|c\alpha + d|} = \frac{k'}{|qa - pc|^r} > \frac{k''}{q^r}$$

since $|qa - pc| < \text{const.}\, q$ for those p/q that approximate $\dfrac{a\alpha + b}{c\alpha + d}$ well. □

Returning to general area-preserving flows it is natural to ask, how to see whether the smoothly conjugate special flow obtained in Proposition 14.2.5 is smoothly conjugate to a suspension. According to the Proposition 2.9.5 a sufficient condition for that is that the base is rotated by a Diophantine angle. Thus we have

Corollary 14.2.7. *Let φ^t be a C^∞ flow of $\mathbb{T}^2$ preserving a C^∞ area element and that possesses a transversal such that the induced map has Diophantine rotation number. Then φ^t is C^∞ conjugate to a linear flow.*

We will see in Section 14.7a how this question can be decided without referring to a section.

Exercises

14.2.1. *Show that for a generic (dense G_δ) set of numbers ρ there exists a real-analytic function φ such that the special flow under φ built over the rotation R_ρ is not C^0 flow equivalent to a linear flow.*

14.2.2. *Prove that every fixed-point-free flow on the Klein bottle has a periodic orbit.*

14.2.3. *Show that any recurrent orbit for a fixed-point-free flow on the Klein bottle is periodic.*

14.2.4. *Prove that for $r \geq 0$ there is a C^∞ flow on the torus that is C^r, but not C^{r+1}, orbit equivalent to a linear flow.*

3. Minimal sets

The examples of minimal sets for flows on surfaces we have seen so far are very limited. Of course fixed points and periodic orbits clearly occur for flows on any compact surface. Furthermore the irrational linear flow on $\mathbb{T}^2$ is minimal. Finally we have mentioned that minimal Cantor sets occur when one builds a special flow over a Denjoy example. The latter is possible for C^1 but impossible for fixed-point-free C^2 flows on the torus. In fact, we will now show that for C^2 flows on any surface the first three examples are the only possible kinds of compact minimal sets. This is a generalization of the Denjoy theorem and the proof again uses a bounded distortion estimate.

Theorem 14.3.1. (Schwartz)[1] *Let M be a smooth surface, φ^t a C^2 flow, and A a nonempty compact minimal set. Then A is either a fixed point or a periodic orbit or $A = M$.*

Remark. Note that if $A = M$ we can conclude that M is compact and φ^t is fixed-point-free, so by the classification of compact surfaces and the Poincaré–Hopf Theorem 8.6.6 M is either the torus or the Klein bottle. Exercise 14.2.3 excludes the latter possibility and thus $M = \mathbb{T}^2$.

As the Denjoy-type flows on $\mathbb{T}^2$ show, the C^2 assumption is really needed. We employ it as follows. A minimal set A has no proper closed invariant subsets, so $\partial A = A$ or $\partial A = \varnothing$. Thus A is nowhere dense unless $A = M$. Thus we need only rule out the possibility of Cantor-type sets, which we accomplish by an argument similar to the proof of the Denjoy Theorem 12.1.1 using the C^2 hypothesis.

Proof. We assume for purposes of contradiction that A is a nowhere dense compact invariant minimal set without fixed or periodic points. Let us take a point in A and consider a C^∞ transverse segment τ through this point whose endpoints are not in A. By taking a flow box $\bigcup_{|t|<\epsilon}\varphi^t(\tau)$ we identify τ with $I=(-\eta,\eta)\subset\mathbb{R}$ and $\tau\cap A$ with a compact nonempty nowhere dense subset C of I. Thus $W:=I\smallsetminus C$ is dense and a countable disjoint union of open intervals (a_l,b_l), $l\in\mathbb{N}$. If U is the set of $x\in I$ whose orbit returns to τ at some time $t>0$ but does not hit an endpoint of τ for time $0<s<t$, then U is a neighborhood of C in I and the return map f is well defined, injective, and continuous, in fact C^2 by transversality, on U. Since τ is transverse to the vector field $\dot{\varphi}^t$ we also have $f'\neq 0$ on U. If V is an open neighborhood of C whose closure is in U then by compactness of $\bar{V}$ there exists $K>1$ such that

$$\frac{1}{K}<|f'|<K,\qquad |f''|<K \tag{14.3.1}$$

on V. Of course C is a nonempty compact invariant minimal set for f without periodic points. Furthermore the set $\tilde{C}:=\bigcup_{l\in\mathbb{N}}\{a_l,b_l\}\smallsetminus\{-\eta,\eta\}\subset C$ of endpoints is f-invariant as well. Thus if (a_l,b_l) is a component of W, $(a_l,b_l)\subset U$, $a_l\neq-\eta$, and $b_l\neq\eta$ then $f((a_l,b_l))$ is also a component of W.

Now consider the distance

$$\epsilon:=d(C,\partial V)\in(0,\eta)$$

and note that the set Q of points $a_l,b_l\in\tilde{C}$ such that $b_l-a_l\geq\epsilon$ is finite. Thus there exists an $N\in\mathbb{N}$ such that $f^n(a_1)\notin Q$ for $n\geq N$. Since $\tilde{C}$ is invariant there exists $l\in\mathbb{N}$ such that $f^n(a_1)=a_l$ or $f^n(a_1)=b_l$. Thus $(a,b):=(a_l,b_l)$ has the property that $f^n((a,b))$ is a component of W of length less than ϵ for all $n\in\mathbb{N}$. This implies that

$$f^n([a,b])\subset V \text{ for } n\in\mathbb{N}$$

since $\{a,b\}\subset C$. These intervals are pairwise disjoint since C contains no periodic points and thus their lengths are summable. We will use this fact now in a Denjoy-type argument which ultimately shows that there must be a periodic point in C after all.

Proposition 14.3.2. *Suppose $f\colon[0,1]\to[0,1]$ is C^2, $K>0$. Then there exists $C\in\mathbb{R}$ such that if $I\subset[0,1]$ and (14.3.1) holds on $\bigcup_{i=0}^k f^i(I)$ then*

$$\Big|\log\frac{(f^{k+1})'(x)}{(f^{k+1})'(y)}\Big|\leq C\sum_{i=0}^k|f^i(x)-f^i(y)|$$

for all $x,y\in I$.

Proof. Since by the chain rule $f^{k+1'}(x) = \prod_{i=1}^{k} f'(f^i(x))$, the Mean Value Theorem yields $\xi_i \subset f^i((p,q))$ such that

$$\begin{aligned}\left|\log\left|\frac{f^{k+1'}(p)}{f^{k+1'}(q)}\right|\right| &\le \sum_{i=0}^{k} |\log|f'(f^i(p))| - \log|f'(f^i(q))|| \\ &\le \sum_{i=0}^{k} |D\log|f'(\xi_i)|| \cdot |f^i(p) - f^i(q)| \\ &= \sum_{i=0}^{k} |f'(\xi_i)|^{-1}|f''(\xi_i)| \cdot |f^i(p) - f^i(q)| \\ &\le K^2 \sum_{i=0}^{k} |f^i(p) - f^i(q)|,\end{aligned}$$

by (14.3.1). □

For later use we note the following immediate consequence:

Lemma 14.3.3. *If $f\colon [0,1] \to [0,1]$ is a C^2 map and I has pairwise disjoint images then the ω-limit set of I contains a critical point.*

Suppose $n \in \mathbb{N}$ and $[p,q] \subset [s,t] \subset V$ such that $f^k([s,t]) \subset V$ for $0 \le k \le n$. Then by Lemma 14.3.3

$$\left|\frac{f^{k+1'}(p)}{f^{k+1'}(q)}\right| \le \exp\left(K^2 \sum_{i=0}^{k} |f^i(p) - f^i(q)|\right). \tag{14.3.2}$$

We first use this observation to show that $|f^{i'}(a)|$ is summable. To that end note that by the Mean Value Theorem there exist $\xi_i \in (a,b)$ such that $f^i(b) - f^i(a) = f^{i'}(\xi_i) \cdot (b-a)$ and hence

$$(b-a)\sum_{i=0}^{\infty} |f^{i'}(\xi_i)| = \sum_{i=0}^{\infty} |f^i(b) - f^i(a)| \le 2\eta$$

since the intervals $f^i((a,b))$ are pairwise disjoint. Since by (14.3.2) we thus have $|f^{i'}(a)| \le |f^{i'}(\xi_i)|e^{2K^2\eta}$, we have shown

$$1 \le d := \sum_{i=0}^{\infty} |f^{i'}(a)| \le \frac{2\eta}{b-a} e^{2K^2\eta}. \tag{14.3.3}$$

We want to see that (14.3.3) implies

$$\lim_{i\to\infty} f^{i'}(x) = 0 \text{ uniformly for } |x-a| \le \delta := \frac{\epsilon}{3K^2 d(1+\eta)} < \frac{\epsilon}{3d} < \epsilon. \tag{14.3.4}$$

To this end we show

Lemma 14.3.4. *For all $n \in \mathbb{N}$*

(1) $f^n([a-\delta, a+\delta]) \subset V$,
(2) $|f^n(x) - f^n(a)| < \epsilon$ *when* $|x-a| \le \delta$,
(3) $|f^{n\prime}(x)| \le e|f^{n\prime}(a)|$ *when* $|x-a| \le \delta$.

Note that (3) indeed implies (14.3.4); (1) and (2) are only used for purposes of induction in the proof.

Proof. We proceed by induction. For $n = 0$ the claim is trivial. Suppose it holds for $k \le n$. We first show (3) for $n+1$. Equation (14.3.2) yields

$$|f^{n+1\prime}(x)| \le |f^{n+1\prime}(a)| \exp\Big(K^2 \sum_{k=0}^{n} |f^k(x) - f^k(a)|\Big)$$

and the Mean Value Theorem together with (3) yields

$$\sum_{k=0}^{n} |f^k(x) - f^k(a)| \le |x-a| \sum_{k=0}^{n} |f^{k\prime}(\xi_k)| \le 3|x-a| \sum_{k=0}^{n} |f^{k\prime}(a)|.$$

By (14.3.3) we get

$$|f^{n+1\prime}(x)| \le |f^{n+1\prime}(a)| e^{3K^2 d\delta} < |f^{n+1\prime}(a)| e^{\frac{\epsilon}{\eta}} < e|f^{n+1\prime}(a)|,$$

proving (3) for $n+1$.

To prove (2) note that by the Mean Value Theorem and (3) we have

$$|f^{n+1}(x) - f^{n+1}(a)| \le |x-a| \cdot |f^{n+1\prime}(\xi)| \le 3\delta |f^{n+1\prime}(a)| \le 3d\delta < \epsilon,$$

proving (2).

Finally, since $f^{n+1}(a) \in C$, (2) and the choice of ϵ imply (1). □

Now we can finish the proof of the theorem. Since C is minimal there exist infinitely many $k \in \mathbb{N}$ such that $|f^k(a) - a| \le \delta/2$. For large enough such k we have $|f^{k\prime}(x)| < 1/2$ for $|x-a| \le \delta$ by (14.3.4) and hence $|\frac{d}{dx}(f^k(x) - x)| > 1/2$, so $f^k(x) - x$ changes sign on $[a-\delta, a+\delta]$ and hence has a zero $z \in [a-\delta, a+\delta]$. We could of course use the Contraction Mapping Principle, Proposition 1.1.2, but in the one-dimensional case the Intermediate Value Theorem is sufficient. Thus $f^k(z) = z$ and furthermore $f^{kn}(a) \xrightarrow[n\to\infty]{} z$ by (14.3.4). Since $a \in C$ and C is closed and invariant we conclude that $z \in C$ contrary to the assumption that C contains no periodic points. This proves the theorem. □

Exercises

14.3.1. *Given an orientable surface of genus g and $1 \le k \le g$ show that there exists a C^1 flow with exactly k nowhere-dense minimal sets that are not fixed points or circles.*

14.3.2. *Show that any C^1 flow on the orientable surface of genus g has no more than g different minimal sets that are not fixed points or periodic orbits.*

4. New phenomena

New types of dynamical behavior appear for smooth flows on the torus with fixed points and for flows on surfaces of higher genus.

a. The Cherry flow.[1] Now we show that in the presence of fixed points behavior that produces Denjoy-type minimal sets on transversals to the flow appears for arbitrarily smooth flows on the torus. The idea is to modify a linear flow in such a way that there is a transversal containing a dense set of points which return to the transversal only finitely may times and are then attracted to a fixed point. The remaining points form a Cantor set which then inevitably exhibits Denjoy-type behavior and consists of points that have a saddle in their orbit closure. This example is then easily modified to give a flow on a surface of genus 2 that has no attracting fixed points and two saddles and exhibits similar phenomena.

Here is a description of this construction. Consider a local vector field on a disk $D_1 \subset \mathbb{R}^2$ with a saddle–node phase portrait as in Figure 7.3.4 and interpolate it via bump functions to the constant vector field $X = (0,1)$ outside a neighborhood D_2 of D_1. Take D_2 of diameter ϵ and rotate the vector field by $\tan^{-1}\alpha$. Translate D_2 to the center of the unit square $[0,1]^2$ and project the restriction to $[0,1]^2$ of this vector field to $\mathbb{T}^2 = [0,1]^2/\mathbb{Z}^2$. This gives a vector field X_0 with a node p and a saddle s in a disk $D \subset \mathbb{T}^2$ which is constant outside D. If ϵ and α are not too big then there will be an interval $[a,b] \subset \{0\} \times S^1$ such that $a, b \in W^s(s)$, $\omega(y) = p$ for $y \in (a,b) \times S^1$ and the return map on $\{0\} \times S^1$ is well defined outside $[a,b]$. Note that this map extends (by constant interpolation across $[a,b]$) to a continuous monotone circle map f_0 of degree one, hence has rotation number $\tau(f_0)$ depending continuously on this construction. Modifying X_0 on $[1-\delta,1] \times S^1$ we can make $Df(x) > 1$ outside $[a,b]$.

Let us show that we can make $\tau(f_0)$ irrational. Let $Y_\lambda = (0, h(x))$ be a C^∞ vector field on $\mathbb{R}^2$ with $\mathrm{supp}(h) \subset [1-\delta, 1]$ such that the map induced between $\{0\} \times \mathbb{R}$ and $\{1\} \times \mathbb{R}$ by the flow of $(\cos\alpha, 0) + Y_\lambda$ is a translation by λ. The vector fields $X_\lambda = X_0 + Y_\lambda$ generate flows on $\mathbb{T}^2$ with induced circle maps f_λ which lift to F_λ on $\mathbb{R}$ such that $\tau(F_1) - \tau(F_0) = 1$, and hence there is a λ_0 for which $\tau(f_{\lambda_0}) \notin \mathbb{Q}$. Let $f := f_{\lambda_0}$.

We call the basin $T = \{q \mid \omega(q) = p\}$ of the sink the *tail* and its complement $\Lambda = \mathbb{T}^2 \smallsetminus T$ the *Cherry set*. To see that $K{:=}\Lambda \cap (\{0\} \times S^1)$ is a Cantor set let $K_0 \subset K$ be a maximal closed interval and $K_n := f^n(K_0)$. Then $l(K_{n+1}) \geq l(K_n)$ and $K_i \cap K_j = \varnothing$ since otherwise we have inclusion by maximality and thus a periodic point of f by Lemma 15.1.2, which is impossible since $\tau(f) \notin \mathbb{Q}$. But then we must have $l(K_0) = 0$ and K has empty interior. Next the pairwise disjoint intervals $I_n := f^{1-n}((a,b))$ have dense union in $\{0\} \times S^1$ and the endpoints belong to different components of $W{:=}W^s(s) \smallsetminus \{s\}$, so each of these components is dense in Λ and $\alpha(x) = \Lambda$ for $x \in W$. This shows that K is perfect, hence a Cantor set.

For the reversed flow the Cherry set is an attractor of a type we have not encountered in the Poincaré–Bendixson setting or in fixed-point-free toral flows.

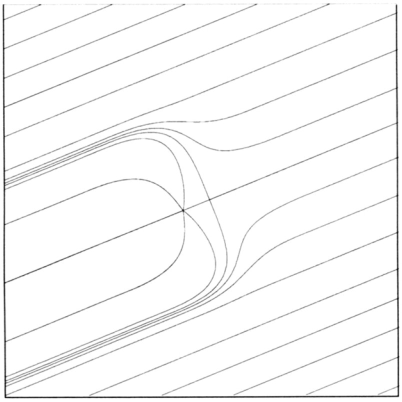

FIGURE 14.4.1. The Cherry flow

It is useful to define a *quasi-minimal set* to be a set containing finitely many fixed points and such that every semiorbit that is not attracted to a fixed point is dense in the set. We can summarize our discussion by saying that the nonwandering set of the Cherry flow consists of an attracting fixed point and a quasi-minimal set containing a hyperbolic saddle. We will see later that quasi-minimal sets are a typical phenomenon for flows on surfaces of higher genus, including area-preserving ones. They produce minimal sets for the Poincaré return map on closed transversals with an appropriately modified topology (see Section 14.5b and Exercise 14.5.2).

The Cherry flow can be modified to produce a flow on the double torus, that is, the sphere with two handles attached: Remove a small neighborhood of the attracting fixed point and consider a second copy of the torus with the same disk removed and the flow reversed. Then these flows can be glued together along the boundaries of the two disks to give a flow on the double torus which has no attracting or repelling fixed points, two saddles, and two disjoint closed invariant nowhere-dense quasi-minimal sets C^+ and C^- containing a saddle each. For every point x outside these two sets the α-limit set is C^- and the ω-limit set is C^+. This flow is obviously not area preserving. Next we will consider an interesting example of an area-preserving flow on the same surface of genus 2.

b. Linear flow on the octagon. If one views the linear flow on the torus as a flow on the unit square whose orbits are parallel, one can naturally try to generalize this construction by replacing the square with another centrally symmetric polygon with opposite sides identified by *translations* and considering the linear flow on the interior extended to the closed surface obtained from the identification. In order to obtain a smooth, or even only continuous flow, certain care has to be taken in defining the flow near the vertices. However, from the point of view of the global orbit structure and the recurrence behavior of the nonfixed points this is not particularly important.

For this kind of construction the next obvious candidate after the square is a regular hexagon. One can see, however, that this construction produces nothing new: The translations of the hexagon tile the plane and the three translations identifying opposite pairs of sides are rationally related, namely, their sum is zero. Thus the group generated by these translations is simply the lattice whose generators are two vectors of equal length at an angle $\pi/3$. This allows us to extend the linear flow to the tiled plane and to view it as a linear flow on the factor by the lattice with two generators, which is again a torus.

To produce a new phenomenon we will consider the next candidate, namely, the regular octagon. It has four pairs of opposite sides which are identified by translations. The translation vectors have equal length and their mutual angles are multiples of $\pi/4$. It is easy to see (cf. Exercise 14.4.3) that the group generated by these translations is not discrete, that is, when we apply the translations to the octagon we will return and cover every point infinitely many times.

When opposite sides of the octagon are identified, all eight vertices are glued together. Notice that from the topological point of view this construction is equivalent to the construction of the genus-two surface from the hyperbolic octagon in Section 5.4e, so the surface thus obtained is homeomorphic to the sphere with two handles. We will give another proof of this fact by constructing a vector field on the surface and calculating its Euler characteristic. This will be the vector field we will later study from the dynamical point of view. Pick a direction in the plane not parallel to any side and take a family of oriented line segments inside the octagon parallel to this direction. The identification of parallel sides allows us to glue those line segments together, except for the ones beginning and ending in a vertex. There are exactly three segments beginning in a vertex and three segments ending in a vertex. A neighborhood of the vertex in the identification space consists of eight sectors glued together in such a way that incoming and outgoing segments alternate and divide the neighborhood into six sectors. No other orbit crosses any of the separating lines. Thus the picture of the orbits near this special point looks similar to the example showed in Figure 8.4.1, except for having six separatrices rather than eight. By making an appropriate time change the flow can be made into a smooth flow with a double saddle.

Now we will give a description of a natural differentiable structure in a neighborhood of the identified vertices which will also provide a natural time change for the flow. The problem is that the Euclidean differentiable structure does not behave well under the projection to the identification space because the total angle is 6π rather than 2π. In the construction on the hyperbolic plane in Section 5.4e the total angle was indeed 2π. In our case the natural way to fix this is to introduce a complex coordinate w on the neighborhoods of the vertex in the octagon such that the standard Euclidean complex coordinate $z = x + iy$ is given by $z - z_0 = w^3$, where z_0 is the coordinate of a vertex. Gluing the edges together with these local coordinates gives a total angle of 2π in the factor. Note that on any open set not containing a vertex the coordi-

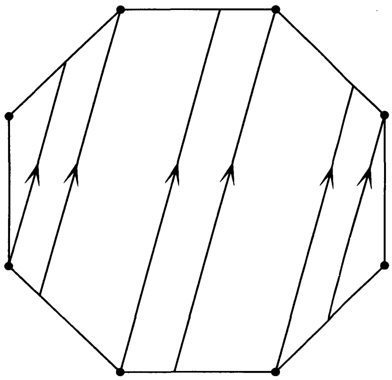

FIGURE 14.4.2. Linear flow with separatrices

nate w is locally differentiably compatible with z, so we have indeed defined a differentiable structure on the surface. We can, in fact, describe the Euclidean area in the coordinate w explicitly. Namely, if $z = x + iy$ then the Euclidean area is given by $dx \wedge dy$. If $w = u + iv$ then $z = w^3 = u^3 - 3uv^2 + i(3u^2v - v^3)$ and thus

$$dx \wedge dy = [(3u^2 - 3v^2)du - 6uvdv] \wedge [6uvdu + (3u^2 - 3v^2)dv] = (3u^2 + 3v^2)^2 dudv.$$

We will now use the fact (which follows from Proposition 5.1.9) that if the flow of a vector field v preserves a volume $\rho\Omega$ then the flow of ρv preserves Ω. Thus in a neighborhood of the vertex we can multiply the vector field by $(3u^2 + 3v^2)^2 = 9|w|^4$ to get a vector field that preserves the standard Euclidean area element of the w-coordinate in the neighborhood. To obtain a vector field on the surface that preserves a smooth area element we multiply by a function ρ that has the following properties: On a small neighborhood of the vertex ρ is equal to the scalar factor we just described. Outside a slightly larger neighborhood we set $\rho = 1$ and interpolate smoothly. The resulting flow preserves Euclidean area outside a neighborhood of the vertex and the w-standard area on a small neighborhood of the vertex. In a small collar around the vertex the invariant area is a smooth multiple of Euclidean area.

Let us now show that the vector field we have thus defined is, in fact, a smooth vector field. Note that the only problem is at the vertex, so we need to check smoothness there. In z coordinates the original vector field is given by a constant vector field $Y = (a, b)$. The coordinate change is given by $z = w^3$, whose derivative is given by $Y = 3w^2X$ or $X = Y/(3w^2)$. Thus the scaled vector field is given by $9|w|^4X = 9w^2\bar{w}^2Y/(3w^2) = 3\bar{w}^2Y$, which is indeed a smooth vector field with a saddle having six separatrices.

Let us now consider the return map to a convenient transversal. As a transversal we take a line connecting the midpoints of two opposite segments of the boundary and such that the angle α between the transversal and the vector field is in $(\pi/4, \pi/2)$. Note that on a neighborhood of such a line the flow-invariant volume is the Euclidean volume, since we are away from the vertex. Consequently the naturally induced volume is just (a constant multiple of) the length element. Consider now the return map to this transversal. It is continuous except at those three points that lie on segments of the flow ending on the saddle. Otherwise the return map is piecewise orientation preserving and, since the invariant volume induces the length element, the image of any interval not containing a point of discontinuity has the same length as the interval itself. Thus the restriction of the return map to every interval without points of discontinuity is a translation and we have a particular case of the situation described in Corollary 14.1.7. Topologically this transversal is a circle and it is the union of the closures of three intervals Δ_1, Δ_2, and Δ_3 without discontinuity points whose lengths are correspondingly $l_1 = 1/(1+\sqrt{2})$, $l_2 = (1+\cot\alpha)/(2+\sqrt{2})$, and $1 - l_1 - l_2$, if we take the transversal to have unit length.

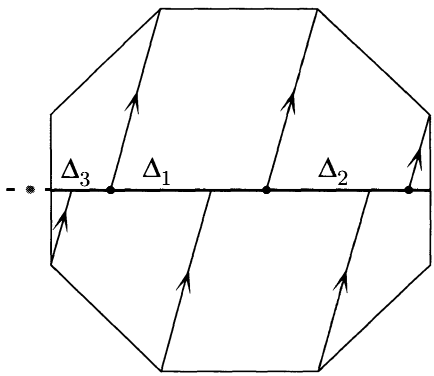

FIGURE 14.4.3. The induced map

The one-parameter family of linear flows we discussed here can also be obtained by considering the billiard problem on a triangular table T with angles $\pi/2$, $\pi/8$, and $3\pi/8$. Denote the vertex with angle $\pi/8$ by v. Note that by reflecting the triangle repeatedly in a side adjacent to the vertex v we obtain an octagon centered at v as a union of 16 copies of T.

Consider an orbit of the billiard in T. Reflecting it in the edges of T, as required, amounts to continuing along a straight line into a reflected copy of T. When the orbit hits the edge of the octagon, reflection in the boundary of T amounts to jumping to the opposite side and continuing in the same direction, that is, the continuation of the orbit corresponds exactly to the translation of the orbit to another copy of the triangle in the opposite orientation. Thus the orbit lifts exactly to an orbit of the linear flow on the octagon of the slope of the billiard orbit. We have thus established a one-to-one correspondence between the collection of orbits of the billiard in the triangle on the one hand and the

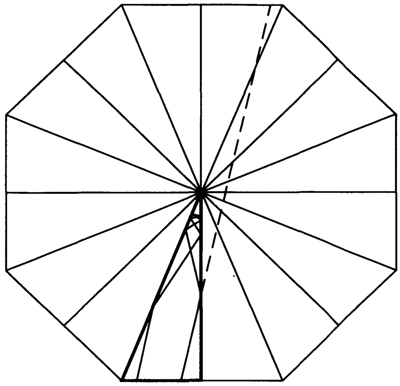

FIGURE 14.4.4. The octagon decomposed

collection of all orbits of the various linear flows on the octagon. It is good to note that the orbits for a fixed linear flow on the octagon correspond to billiard orbits whose angles with respect to a fixed side take at most 16 values, namely, for a given angle α we obtain the angles $\pm\alpha + k\pi/4$, that is, the angles obtained by reflecting α with the various reflections we used to generate the octagon.

We will soon prove a general result (Corollary 14.6.6) which implies that for all but countably many directions the linear flow on the octagon is topologically transitive and quasi-minimal. This is similar to the situation for linear flows on the torus (see Proposition 1.5.1).

Exercises

14.4.1. *Given an orientable surface of genus g and $1 \leq k \leq g$ show that there exists a C^∞ flow with exactly k nowhere-dense quasi-minimal sets that are not circles.*

14.4.2. *Show that any C^∞ flow on the orientable surface of genus g has no more than g different quasi-minimal sets that are not fixed points or circles.*

14.4.3. *Prove that the orbit of any point with respect to the group generated by translations identifying opposite sides of the regular octagon is dense.*

14.4.4. *For the linear flow on the octagon consider the transversal given by a diagonal (a diameter) of the octagon. Discuss the map on it induced by the flow by describing the topology of the transversal, giving the lengths of the maximal intervals without discontinuities and the way in which they are permuted.*

14.4.5. *Generalize the construction of Subsection b to the regular $2n$-gon ($n \geq 4$). Calculate the genus of the resulting surface and the number and indices of the fixed points for the flow, and describe the differentiable structure that makes a time change smooth.*

5. Interval exchange transformations

a. Definitions and rigid intervals. The return map for the linear flow on the octagon is an example of the maps that appear as section maps of area-preserving flows on surfaces with finitely many saddles (see Corollary 14.1.7).

Definition 14.5.1. Consider a permutation π of $\{1,\dots,n\}$, a vector $v = (v_1,\dots,v_n)$ in the interior of the unit simplex, that is, such that $v_i > 0$ for $i = 1,\dots,n$ and $\sum_{i=1}^n v_i = 1$, and a vector $\epsilon = (\epsilon_1,\dots,\epsilon_n)$ whose coordinates are either 1 or -1. Let $u_0 = 0$, $u_i = v_1+\cdots+v_i$ for $i = 1,\dots,n$, $\Delta_i = (u_{i-1}, u_i)$ for $i = 1,\dots,n$. The *interval exchange transformation* $I_{v,\pi,\epsilon}\colon [0,1] \to [0,1]$ is the map that is continuous and Lebesgue-measure preserving on every interval Δ_i, rearranges those intervals according to the permutation π, and preserves or reverses orientation on Δ_i according to the sign of ϵ_i $(i = 1,\dots,n)$. If $\epsilon_i = 1$ for all i we write $I_{v,\pi}$ instead of $I_{v,\pi,\epsilon}$. Such a map will be called an *oriented interval exchange transformation.*

Remark. One can similarly define an arc exchange transformation of the circle. Obviously every arc exchange transformation of n arcs is also an interval exchange transformation of $n+1$ intervals.

Thus the map $I_{v,\pi,\epsilon}$ restricted to each interval Δ_i is either a translation of Δ_i (if $\epsilon_i = 1$) or a reflection with respect to a point (if $\epsilon_i = -1$).

The return map to the midpoint transversal for the linear flow on the octagon with slope α with respect to the horizontal is an interval exchange of four intervals (obtained from an exchange transformation of three arcs of a circle).

While talking about interval exchange transformations it is convenient to use the term "partition" in an extended sense as a decomposition of $[0,1]$ or an interval $\Delta \subset [0,1]$ into intervals with piecewise disjoint interiors. If two successive interiors Δ_i and Δ_{i+1} are mapped to $\Delta_k \cup \Delta_{k+1}$ preserving orientation or to $\Delta_k \cup \Delta_{k-1}$ reversing orientation, then we can lump $\Delta_i \cup \Delta_{i+1}$ together and consider $I_{v,\pi,\epsilon}$ as an exchange of a smaller number of intervals. Thus, without loss of generality we may always assume that $u_1,\dots,u_{n-1}$ are discontinuity points of I and that the partition $\xi := \xi(I)$ is the partition into intervals of continuity of I.

There is an ambiguity in the definition of the map $I_{v,\pi,\epsilon}$ at the points of discontinuity, that is, at $u_1, u_2,\dots,u_{n-1}$. Sometimes there is a natural way to extend the definition to some of those points and obtain a one-to-one map. For example, for $n = 2$, $\pi = (2,1)$, there is only one discontinuity point v_1 inside the interval and if we set $I_{v,\pi}(v_1) = 0$ then by identifying 0 and 1 we obtain the rotation of the circle by the angle $2\pi v_2$. More often, however, such a natural extension is not possible as in the octagon example in Section 4b. A more useful approach is the following. At each point u_i of discontinuity the map $I_{v,\pi,\epsilon}$ has left and right limits, which we will denote by w_i^- and w_i^+, correspondingly. It makes sense to think of the point u_i as having two "ends" and of w_i^- and w_i^+ as the images of those "ends".

Definition 14.5.2. An interval exchange transformation $I = I_{v,\pi,\epsilon}$ is said to have a *saddle connection* if for some $i, j \in \{0, \dots, n\}$ and for some $k \in \mathbb{N}$ we have $\bar{I}^k(w_i^+) = u_j$ (correspondingly, $I^k(w_i^-) = u_j$) but $I^l(w_i^+)$ (correspondingly, $I^l(w_i^-)$) are points of continuity for $0 < l < k$. The orbit segment $u_i, w_i^+, I(w_i^+), \dots, I^{k-1}(w_i^+), u_j$ (correspondingly, $u_i, w_i^-, \dots, I^{k-1}(w_i^-), u_j$) is called a *connecting segment*. Let us call a permutation π of $\{1, \dots, n\}$ *irreducible* if it does not preserve any subset of the form $\{1, \dots, k\}$ for $k = 1, \dots, n-1$. An open interval $\Delta \subset [0, 1]$ is called *rigid* under I if all positive iterates of I are defined and continuous on Δ. A rigid interval Δ is a *maximal* rigid interval if any other rigid interval is either disjoint from it or contained in it. An interval exchange transformation is called *generic* if it has no rigid intervals. A point $x \in [0, 1]$ is called a *generic point* for I if all positive and negative iterates of x are defined, that is, no image or preimage of x is a discontinuity point.

Evidently we have

Lemma 14.5.3. *The number of different connecting segments for any interval exchange transformation $I_{v,\pi,\epsilon}$ does not exceed $2n - 2$, where n is the number of intervals of continuity.*

Lemma 14.5.4. *Any rigid interval Δ for an interval exchange transformation I consists of periodic points. Any maximal rigid interval either consists of points of the same period or all of its points except for the midpoint have even period $2k$ while the midpoint has period k. Any endpoint of a maximal rigid interval belongs to a connecting segment.*

Proof. It suffices to consider maximal rigid intervals since the union of rigid intervals intersecting a given rigid interval Δ is again a rigid interval and clearly maximal.

For any rigid interval Δ any image $I^k(\Delta)$ is an interval of the same length and if Δ is maximal then $I^k(\Delta) \cap \Delta$ is either empty or equal to Δ. If $I^k(\Delta) \cap \Delta = \varnothing$ for all k then $I^k(\Delta) \cap I^l(\Delta) = \varnothing$ for all $k, l \geq 0$ by invertibility and continuity of I on every image of Δ. But this is impossible since the sum of the lengths of these intervals is infinite. Thus there exists $k \in \mathbb{N}$ such that $I^k(\Delta) = \Delta$. Since I^k preserves Lebesgue measure and is continuous on Δ, it is either the identity transformation or the reflection in the midpoint, which is thus fixed, and I^{2k} is the identity on Δ.

Let x be the left endpoint of Δ. I^k (correspondingly, I^{2k}) must be discontinuous at x because otherwise it would be the identity on a neighborhood of x, contradicting maximality of Δ. The same argument applies to I^{-k} (correspondingly, I^{-2k}). Since discontinuity appears only when the image of a point is one of the points $u_1, \dots, u_{n-1}$, we see that x belongs to a connecting segment. The same argument applies to the right endpoint of Δ. □

Thus we can associate with any maximal rigid interval its *orbit*, that is, the union of its images. Since for an exchange transformation of n intervals the number of different connecting segments does not exceed $2n - 2$ and the same connecting segment can only be an endpoint for one maximal rigid interval, there are at most $2n-2$ maximal rigid intervals whose orbits are different.

Corollary 14.5.5. *If an interval exchange transformation has no saddle connections then it is generic.*

Corollary 14.5.6. *On any rigid interval Δ all (positive and negative) iterates of I are defined and continuous.*

The joint partition $\xi^I_{-m}:=\xi\vee I^{-1}(\xi)\vee\cdots\vee I^{1-m}(\xi)$ into intervals of continuity of I^k, $k=1,\dots,m$, consists of at most $m(n-1)+1$ intervals and exactly this many if there are no saddle connections. For the nested sequence ξ^I_{-m}, $m\in\mathbb{N}$, there is an obvious dichotomy:

(1) $\max_{c\in\xi^I_{-m}} l(c)\to 0$ as $m\to\infty$, where $l(\cdot)$ is the length. This happens for generic interval exchange transformations.
(2) There is a nested sequence of elements $c_m\in\xi^I_{-m}$ such that $c_\infty := \bigcap_{m=1}^{\infty} c_m$ is an interval of positive length. Then c_∞ is the closure of a maximal rigid interval and obviously any maximal rigid interval appears in this way.

b. Coding. Let L be the union of the closures of all maximal rigid intervals. The partition ξ offers a natural way of coding the interval exchange transformation on the invariant set $[0,1]\smallsetminus L$. Namely, let

$$\Omega_I := \Big\{\omega\in\Omega_n \mid \bigcap_{m\in\mathbb{Z}} I^{-m}(\Delta_{\omega_m+1})\neq\varnothing\Big\}. \tag{14.5.1}$$

Ω_I is closed (Exercise 14.5.1) and obviously shift invariant. If I is generic then the map $h\colon\Omega_I\to[0,1]$, $\omega\mapsto\bigcap_{m\in\mathbb{Z}} I^{-m}(\Delta_{\omega_m+1})$ is a finite-to-one continuous surjective map and injective on preimages of generic points. If I has no saddle connection then any nongeneric point has exactly two preimages. Otherwise 2^n is obviously an upper bound for the number of preimages of a point.

If I has rigid intervals then h is not defined on Ω_I. However, it is still defined and continuous on the complement of finitely many periodic orbits corresponding to the orbits of maximal rigid intervals, and its image is $[0,1]\smallsetminus L$. Even though this map cannot be extended to a semiconjugacy between a symbolic system and I, it is a measure-theoretic isomorphism for any nonatomic ergodic shift-invariant measure on Ω_I since both the set of discontinuity and the set of nonuniqueness can only carry atomic ergodic measures.

Notice the analogy to the measure-theoretic classification of circle maps (Theorem 11.2.9) in the nontransitive case.

c. Structure of orbit closures. We will use notions from topological dynamics, such as recurrence, freely in the sequel. Although these were defined for continuous maps this is justified because on one hand these definitions do not require continuity and we do not use properties that do, and on the other hand because we have a symbolic model at hand, where all these notions appear in the standard way.

The construction of the first-return map plays a very important role in the study of interval exchange transformations. This is due to the following lemma which shows that the class of interval exchange transformations is closed under the operation of inducing (taking the return map) on subintervals and the number of intervals increases by at most 2.

Lemma 14.5.7. *Let I be an exchange transformation of n intervals or arcs of the circle. Then for any interval (or arc) Δ the first-return map I_Δ is defined and continuous everywhere except for at most $n+1$ points and is an exchange transformation of $k \le n+2$ intervals.*

Proof. By the Poincaré Recurrence Theorem 4.1.19 the set of points that return to Δ has full Lebesgue measure and is hence dense. Suppose $x \in \Delta$ is such that $y = I_\Delta(x) = I^k(x) \in \mathrm{Int}(\Delta)$ and $I^l(x)$ $(0 \le l < k)$ are points of continuity of I. Then I^k is a local isometry near x and hence maps a neighborhood of x onto a neighborhood of y in Δ. On the other hand, $\min_{1 \le l < k} \mathrm{dist}(I^l(x), \Delta) = \epsilon > 0$, so $I_\Delta = I^k$ in a neighborhood of x. Thus I_Δ is continuous at x, hence I_Δ is defined and continuous on an open set.

Let z be the left endpoint of a maximal interval where I_Δ is defined and continuous. This means that $I_\Delta = I^k$ in a one-sided neighborhood of z for some k. But if $I^l(z)$, $0 \le l \le k$ are points of continuity of I and not endpoints of Δ then $I^k(z) \in \mathrm{Int}\,\Delta$ and by the previous argument $I_\Delta = I^k$ in a two-sided neighborhood of z, a contradiction. Thus an iterate $I^l(z)$ for $0 \le l \le k$ is either a point of discontinuity of I or an endpoint of Δ. Consider the smallest such l. Each of the $n-1$ points of discontinuity of I and each of the two endpoints of Δ can appear in this way as an iterate of at most one left endpoint of an interval of continuity of I_Δ. Thus I_Δ is defined away from at most $n+1$ points in $\mathrm{Int}\,\Delta$ and is an isometry on the complementary intervals. □

Remark. The above argument essentially reproves Proposition 14.1.6 in the setting of interval exchange transformations.

Corollary 14.5.8. *Every generic point for an interval exchange transformation is recurrent.*

Now we can prove our first important finiteness result for interval exchange transformations.

Proposition 14.5.9. *Let x be a nonperiodic recurrent point for an interval exchange transformation I. Then the complement of the orbit closure of x consists of finitely many intervals whose endpoints belong to connecting segments.*

Proof. Let $\Delta = (a, b)$ be such a complementary interval. By Lemma 14.5.7 I_Δ is defined at all but finitely many points of Δ, so the right-hand limit $I_\Delta(a^+)$ at a of $I_\Delta(x)$ is well defined. There are two possibilities:

(1) $I_\Delta(a^+) \in \operatorname{Int}\Delta$. Then $I_\Delta(a^+) = I^k(a^+)$ for some minimal $k > 0$. If I^k is continuous at a then I^k is an isometry near a and hence points from the orbit of x that accumulate at a are mapped to points accumulating on $I^k(a) \in \operatorname{Int}\Delta$, a contradiction. Thus $I^l(a)$ is a discontinuity point of I for some $l < k$.

(2) $I_\Delta(a) \in \partial\Delta$, that is, $I_\Delta(a^+) = a$ or $I_\Delta(a^+) = b$. In the first case $I_\Delta = I^k = \operatorname{Id}$ in a right-hand neighborhood of a. If a is a continuity point of I^k then $I^k = \operatorname{Id}$ on a neighborhood of a contrary to the fact that nonperiodic recurrent orbits accumulate on a; hence an iterate of a is a discontinuity point of I. The case $I_\Delta = I(a^+) = b$ reduces to the previous one or (1) by considering I_Δ^2.

Thus we find in all cases that the positive semiorbit of a contains a discontinuity point of I. The same argument and conclusion apply for I^{-1}, so a belongs to a connecting segment. By symmetry the same holds for b. □

Corollary 14.5.10. *Every generic orbit is either periodic or its orbit closure is a finite union of intervals.*

Corollary 14.5.11. *For a generic interval exchange transformation the closure of all but finitely many orbits is a union U of finitely many intervals. The orbit of any generic point in U is dense in U.*

Corollary 14.5.12. *If an interval exchange transformation does not have saddle connections then every generic orbit as well as every semiorbit that does not contain a discontinuity point is dense.*

Note that the latter corollary describes a situation as close to topologically minimal as it could be for an interval exchange. In fact, the symbolic model Ω_I is a minimal set in this case (Exercise 14.5.2).

Let us call the orbit closures consisting of finitely many intervals (of nonzero length) the *transitive components* of the interval exchange transformation I. The interiors of different transitive components are disjoint. Similarly let us call the orbit of a maximal rigid interval a *periodic component*. We thus have found that all points, except maybe those lying on connecting segments, have to belong either to a transitive or to a periodic component. It follows from Lemma 14.5.4 and Proposition 14.5.9 that the boundary of each (transitive or periodic) component consists of complete connecting segments. The number of connecting segments does not exceed $2n - 2$. Each connecting segment may belong to the boundary of at most 2 components. Thus the total number of components does not exceed $4n - 4$. Furthermore in the oriented case each connecting segment comes with two orientations and the boundary of each component must contain at least one positively oriented and one negatively oriented segment, which decreases the possible number of components to $2n-2$. Thus the topological structure of orbits of interval exchange transformations can be summarized as follows:

Theorem 14.5.13. *Let I be an exchange map of n intervals. Then $[0,1]$ splits into a finite union of connecting segments and $k \leq 4n-4$ disjoint open invariant sets each of which is either a transitive or a periodic component and is a finite union of open intervals. If in addition I is oriented then $k \leq 2n-2$.*

d. Invariant measures. The last theorem gives a topological finiteness result for interval exchange transformations. In fact, every transitive component is *quasi-minimal*: Any semiorbit that does not begin or end in a discontinuity point is dense in it. A corresponding measure-theoretic property would be uniqueness of the nonatomic invariant measure on a transitive component. We shall see in the next subsection that this is not the case even under the stronger assumption of no saddle connections. Nevertheless there is the following general finiteness result for invariant measures:

Theorem 14.5.14. *Let I be an exchange transformation of n intervals. Then there are at most n mutually singular invariant nonatomic Borel probability measures for I supported on the union of the transitive components.*[1]

Corollary 14.5.15. *There are at most n distinct invariant nonatomic ergodic Borel probability measures for I.*

Corollary 14.5.16. *For a generic interval exchange there are no more than n disjoint invariant sets of positive Lebesgue measure.*

Proof. Since such a measure is supported on the transitive components, the joint partitions constructed from ξ are dense with respect to the metric $\mathcal{D}$ of (4.3.9), so ξ is a one-sided generator. (See Section 4.3 for a general discussion.) Thus an invariant measure is determined by its values on the elements of the joint partitions ξ_m^I. The key observation is that it is indeed determined by its values on the intervals in ξ. Namely, these determine the measures of the elements of the joint partition $\xi \vee I(\xi)$ as follows: Start from the left endpoint and notice that the first interval of $\xi \vee I(\xi)$ is the shorter of the leftmost interval of ξ and the leftmost image of an interval in ξ, so its measure is determined. The next interval is again the shorter of the remainder of the other interval on the left and the next image, and so forth. Similarly we can proceed by induction to determine from the values of the measure on the intervals of $\xi \vee \dots \vee I^n(\xi)$ and ξ those on the intervals of $\xi \vee \dots \vee I^{n+1}$ by superimposing $I(\xi \vee \dots \vee I^n(\xi))$ on ξ. This defines a map h from the I-invariant measures to the $(n-1)$-dimensional simplex σ in $\mathbb{R}^n$ which is evidently affine, continuous (in the weak* topology), and, as we saw, injective. Let us note that mutually singular measures correspond to linearly independent elements of σ. Namely, if $a_1 h(\mu_1) + \dots + a_l h(\mu_l) = 0$ then on a set A with $\mu_i(A) > 0$ and $\mu_j(A) = 0$ for $i \neq j$ we have $0 = a_1\mu_1(A) + \dots + a_l\mu_l(A) = a_i\mu_i(A)$, whence $a_i = 0$. □

Remark. The image of this set of measures is always a simplex and its vertices correspond to ergodic measures.

Let us notice that different invariant measures for an interval exchange generate conjugacies between this and other interval exchanges. To facilitate this discussion assume the interval exchange is topologically transitive, which is weaker than absence of saddle connections but stronger than being generic. In this case every nonatomic invariant measure is positive on open sets, and hence the map taking an interval $[0,t]$ to its measure is a homeomorphism and takes this interval exchange to another one such that the image of the given invariant measure is Lebesgue measure. Thus, if a topologically transitive interval exchange has k ergodic invariant measures then we have a $(k-1)$-simplex of topologically conjugate interval exchanges.

e. Minimal nonuniquely ergodic interval exchanges. We will soon see that the absence of saddle connections which (by Corollary 14.5.12) implies essential minimality, is a typical property in many families of interval exchange transformations. We now give an example of an interval exchange transformation that has no saddle connections and is not uniquely ergodic. Both the method of construction and the result are very similar to those that appear in Corollary 12.6.4. Suppose $s\colon [0,1] \to \{0,1\} = \mathbb{Z}/2$ has three points of discontinuity. Then the extension

$$I_s\colon [0,1] \times \{0,1\} \to [0,1] \times \{0,1\}, \quad I_s(x,i) := (I(x), i + s(x))$$

of an exchange I of m intervals can be viewed as an exchange of the at most $2(m+3)$ intervals obtained from the two copies of the (at most $m+3$) intervals defined by subdividing the intervals of continuity of I, if necessary, at the discontinuity points of s. Our example then becomes a corollary of the following coboundary construction:

Theorem 14.5.17. *Let I be an oriented interval exchange transformation without saddle connections. Then there is a function $s\colon [0,1] \to \{0,1\}$ with three discontinuity points such that*

(i) *$s(x) = h(I(x)) - h(x)$ for some measurable $h\colon [0,1] \to \{0,1\}$, and*
(ii) *for $O \subset [0,1]$ open $\lambda(h^{-1}(\{1\}) \cap O) > 0$, $\lambda(h^{-1}(\{0\}) \cap O) > 0$.*[2]

Thus h is metrically dense (Definition 12.6.2) and s is exactly a wild coboundary, similarly to Proposition 12.6.3. Thus, similarly to Corollary 12.6.4 we obtain our desired example:

Corollary 14.5.18. *The interval exchange given by the extension I_s corresponding to s has no saddle connections and is not ergodic.*

Proof. By (i) of Theorem 14.5.17, s is a measurable coboundary and hence I_s is metrically isomorphic to $I \times \mathrm{Id}$ via $(x,i) \mapsto (x, i + h(x))$, so I_s preserves graph h and graph$(1-h)$, both of positive (product) measure. On the other hand we can see that there are no saddle connections by considering a point w_i^- (as before Definition 14.5.2) and noting that its positive I-semiorbit is dense by Corollary 14.5.12, so by (ii) of Theorem 14.5.17 it is dense for I_s and hence not part of a connecting segment. Points w_i^+ are taken care of similarly and evidently no points outside the positive semiorbit of the $w_i^{\pm}$ could be part of a connecting segment. □

Proof of Theorem 14.5.17. We will use addition modulo 1, that is, we are allowed to switch + and − signs. Replacing I by I^{-1} we may replace (i) by

(i′) $s(x) = h(I^{-1}(x)) - h(x)$.

We call an interval $\Delta \subset [0,1]$ *k-clear* if I^i is continuous on $\operatorname{Int}\Delta$ for $|i| \le k$. For $n \in \mathbb{N}_0$ let $a_n := I^n(0)$. Due to the absence of saddle connections this sequence is well defined and by Corollary 14.5.12 dense in $[0,1]$. For $a_n < a_m$ let $\chi_{m,n} = \chi_{[a_n,a_m)}$ be the characteristic function.

We now begin an inductive construction similar to that of Proposition 12.6.3, although more explicit. First let $k_0 = 0$ and $k_1 > 0$ be such that $[a_0, a_{k_1}] \cap [a_1, a_{k_1+1}] = \varnothing$. Inductively we will determine an increasing sequence $\{k_m\}_{m\in\mathbb{N}_0}$ and a related sequence l_m defined by

$$l_0 = 1, \quad l_1 = k_1+1, \quad l_{m+1} = k_{m+1}-k_m+l_{m-1}, \text{ that is, } l_m = 1+\sum_{i=0}^{m-1}(-1)^i k_{m-i}$$

so that

(1) $a_{k_0} < \cdots < a_{k_m} < a_{l_0} < a_{l_2} < \cdots < a_{l_{2[m/2]}} < a_{l_1} < \cdots < a_{l_{2[(m+1)/2]-1}}$,
(2) $[a_{k_m}, a_{k_{m+1}}]$ is k_m-clear,
(3) $a_{k_{m+1}} - a_{k_m} < (a_{k_m} - a_{k_{m-1}})/3k_m$.

To see that (1)–(3) can be satisfied inductively assume they hold up to m. By (1) there is a c such that $a_{k_m} < c < a_{l_0}$ and $[a_{k_m}, c)$ is k_m-clear. By density of $\{a_n\}$ there exists $k_{m+1} > k_m$ such that $a_{k_{m+1}}$ is in the left half of $[a_{k_m}, c)$, showing (2) for $m+1$. Taking $a_{k_{m+1}}$ still closer to a_{k_m} also yields (3). To verify (1) note that $a < \cdots < a_{k_{m+1}} < a_{l_0}$ is already known. Next note (by induction) that $l_m = k_m - l_{m-1} + 2 \le k_m + 1 \le k_{m+1}$ and hence $-k_m < l_{m-1} - k_m \le 0$, so (2) implies that $[a_{l_{m-1}}, a_{l_{m+1}}] = I^{l_{m-1}-k_m}([a_{k_m}, a_{k_{m+1}}])$ is a translate of $[a_{k_m}, a_{k_{m+1}}]$, which yields the remaining inequalities in (1).

To construct s let $s_1(x) := \chi_{0,k_1}(x) + \chi_{1,k_1+1}(x) = \chi_{0,k_1}(x) - \chi_{0,k_1}(I^{-1}(x))$ and

$$\begin{aligned} s_{m+1}(x) :=& s_m(x) + \chi_{k_m,k_{m+1}}(x) - \chi_{l_{m-1},l_{m+1}}(x) \\ =& s_m(x) + \chi_{k_m,k_{m+1}}(x) - \chi_{k_m,k_{m+1}}(I^{k_{m+1}-l_{m+1}}(x)) \\ =& s_m(x) + g_{m+1}(I^{-1}(x)) - g_{m+1}(x), \end{aligned}$$

where $g_{m+1}(x) = \sum_{i=1}^{k_m-l_{m-1}} \chi_{k_m-i,k_{m+1}-i}(x)$. With (1) this yields

$$\begin{aligned} s_m(x) &= \sum_{i=1}^{m} \chi_{k_{i-1},k_i}(x) + \chi_{l_0,l_1}(x) - \sum_{i=1}^{m-1} \chi_{l_{i-1},l_{i+1}}(x) \\ &= \chi_{0,k_m}(x) + \chi_{l_{2[m/2]},l_{2[(m+1)/2]-1}}(x) \to \chi_{[0,b)}(x) + \chi_{[c,d)}(x) =: s(x) \end{aligned}$$

(note that we add mod 2), where $b = \lim_{m\to\infty} a_{k_m}$, $c = \lim_{m\to\infty} a_{l_{2m}}$, $d = \lim_{m\to\infty} a_{l_{2m+1}}$. Thus s has three points of discontinuity. Let us show that s is a coboundary. First, $c_m(x) = h_m(I^{-1}(x)) - h_m(x)$, where

$$h_m(x) := \chi_{0,k_1}(x) + \sum_{i=2}^{m} g_i(x) = \chi_{0,k_1}(x) + \sum_{j=2}^{m} \sum_{i=2}^{k_{j-1}-l_{j-2}} \chi_{k_{j-1}-i,k_j-i}(x).$$

Since $\lambda(g_{m+1}^{-1}(\{1\})) \le (k_m - l_{m-1})(a_{k_{m+1}} - a_{k_m}) < (a_{k_m} - a_{k_{m-1}})/3$ by (3), h_m converges in L^1 to a function h which clearly satisfies (i′).

Next we prove (ii). We call $[a_{k_{m-1}+i}, a_{k_m+i})$ an interval of rank m if $i \in \{0,\dots,k_{m-1} - l_{m-2} - 1\}$. Such intervals are either disjoint or there is an inclusion: Suppose $n \le m$, $i \in \{0,\dots,k_{m-1} - l_{m-2} - 1\}$, $j \in \{0,\dots,k_{n-1} - l_{n-2} - 1\}$, and $a_{k_{m-1}+i} \le a_{k_{n-1}+j} < a_{k_m+i}$. Since $k_{n-1} + j < 2k_{n-1} \le 2k_{m-1}$, applying $I^{-(k_{n-1}+j)}$ gives $a_{k_{m-1}+(j-k_{n-1}-j)} \le 0 < a_{k_m+(j-k_{n-1}-j)}$, so we have equality on the left and hence $a_{k_{m-1}+i} = a_{k_{n-1}+j}$. In particular intervals of a given rank are pairwise disjoint and h_m is constant on every interval of rank m.

For any interval $\Delta \subset [0,1]$ there is an interval $\Delta' \subset \Delta$ of rank m for some m by density of $\{a_m\}$. h_m is constant on Δ' and for $n > m$ (3) yields

$$\begin{aligned}\lambda(\{x \in \Delta' \mid h_n(x) = h_m(x)\}) &\ge \lambda(\Delta') - \sum_{i=1}^{n-m} \lambda(g_{n+i}^{-1}(\{1\})) \\ &\ge \lambda(\Delta')\Big(1 - \frac{1}{3} - \frac{1}{9} - \cdots - \frac{1}{3^{n-m}}\Big) \ge \lambda(\Delta')/3\end{aligned}$$

and $\lambda(\{x \in \Delta' \mid h(x) = h_m(x)\}) \ge \lambda(\Delta')/3$. On the other hand for the smallest $m' > m$ such that Δ' contains an interval Δ'' of rank m' the constant value of $h_{m'}$ on Δ'' differs from that of h_m on Δ' while the same argument as before shows that $\lambda(\{x \in \Delta'' \mid h(x) = h_{m'}(x)\}) \ge \lambda(\Delta'')/3$. □

Exercises

14.5.1. *Show that the set Ω_I defined in (14.5.1) is closed in Ω_n.*

14.5.2. *Show that if I has no saddle connections then the shift on Ω_I is minimal.*

14.5.3. *Prove that every interval exchange has zero entropy with respect to any invariant measure.*

14.5.4. *Consider the permutation $\pi = (3,2,1)$. Show that the oriented interval exchange transformation $I_{v,\pi}$ for any vector v can be obtained by inducing a circle rotation on an interval. Prove that in this case minimality implies unique ergodicity and find a necessary and sufficient condition for minimality.*

14.5.5. *Consider an exchange of two arcs on the circle that changes orientation on one arc. Prove that all orbits are periodic.*

14.5.6. *Consider the map* $I \times \mathbb{Z}_2$, $(x, j) \mapsto (x + \alpha, j\chi_\beta(x))$, *where* χ_β *is the characteristic function of* $[0, \beta]$. *Show that there exist* α, β *such that this map has no saddle connection but is not uniquely ergodic.*

6. Application to flows and billiards

a. Classification of orbits. Corollary 14.1.7 shows that interval exchange transformations appear as (or, more precisely, are smoothly conjugate to) Poincaré maps induced by area-preserving flows on compact surfaces with fixed points of the saddle type. There is also a more general case where the theory of interval exchange transformations applies.

Let $f: [0, 1] \to [0, 1]$ be a piecewise-monotone map that is one-to-one and continuous away from finitely many points. A convenient way to represent such a transformation is as $f = I \circ h$, where h is a homeomorphism and I an interval exchange transformation. Let μ be a nonatomic f-invariant Borel probability measure. Then the map $g: [0, 1] \to [0, 1]$, $g(x) := \mu([0, x])$ is monotone and defines a semiconjugacy between f and an interval exchange transformation, for the factor map preserves Lebesgue measure and has only finitely many discontinuity points. There is an obvious but important case where the semiconjugacy in fact turns out to be a conjugacy.

Proposition 14.6.1. *Any* $f = I \circ h$ *as above that preserves a measure positive on open intervals is topologically conjugate to an interval exchange transformation.*

Using Poincaré maps for flows yields the following result.

Proposition 14.6.2. *Let* φ *be a* C^0 *flow on a closed compact surface defined by a uniquely integrable* C^0 *vector field* X, *and* τ *a transversal to* X. *Suppose that* φ *has a finite number of fixed points, which are orbit equivalent to (multiple) saddles or centers. Furthermore suppose that* φ *preserves a Borel probability measure that is positive on open sets. Then the first-return map induced on* τ *is topologically conjugate to an interval exchange transformation.*

Proof. By the remark after Proposition 14.1.6 the return map is defined away from finitely many points, which are the last points of intersection of the stable separatrices of a saddle with τ. Assume τ does not pass through a center. The return map preserves the measure ν defined by $\nu(A) := \mu(\bigcup_{t=0}^{\epsilon} \varphi^t(A))/\epsilon$ for any $\epsilon > 0$ that is smaller than the minimal return time for τ. Since $\nu(A) > 0$ for any open $A \subset \tau$, Proposition 14.6.1 applies. □

This result allows us to apply Theorem 14.5.13 to measure-preserving flows on surfaces.

Theorem 14.6.3. *Under the assumptions of Proposition 14.6.2 the surface M splits in a φ-invariant way as $M = \bigcup_{i=1}^{k} P_i \cup \bigcup_{j=1}^{l} T_j \cup C$, where l is at most equal to the genus of M, P_i are open sets consisting of periodic orbits, each T_j is open, every semiorbit in T_j that is not an incoming separatrix of a fixed point is dense in T_j, and C is a finite union of fixed points and saddle connections.*[1]

Corollary 14.6.4. *If in addition φ has no saddle connections it is quasi-minimal, that is, every semiorbit other than the fixed points and separatrices is dense in M.*

Remark. Similarly to the previous section we will call the P_i *periodic components* and the T_j *transitive components* of the flow.

Proof. One can construct a finite family of closed transversals that intersects every semiorbit of φ except for fixed points. Now apply Theorem 14.5.13 to the return map on each transversal and take the images of each periodic and transitive component under the flow. They produce the sets P_i and T_j. Any orbit not included in these sets must be a saddle connection for the flow. It remains to show that $l \leq \text{genus}(M)$. First pick an orbit segment in each of the T_j that almost returns back and close it up to obtain a pretransversal γ_j. Preservation of measure implies that $M \smallsetminus \gamma_j$ is connected. Since the γ_j are disjoint the same argument applies to $M \smallsetminus \bigcup_{j=1}^{l} \gamma_j$, so $l \leq \text{genus}(M)$ (see the remarks after Theorem A.5.2). □

Remark. In the absence of centers one can give estimates on $k+l$. An easy one involves counting the number of incoming separatrices using the Poincaré–Hopf Index Formula (Theorem 8.6.6); a more subtle argument gives $k+l \leq \text{genus}(M)$.

b. Parallel flows and billiards in polygons. The case of best recurrence properties appears when the decomposition of Theorem 14.6.3 contains a single transitive component and no periodic component, that is, when M is a quasi-minimal set for the flow as defined at the end of Section 14.4a. By Corollary 14.6.4 the absence of saddle connections is sufficient for quasi-minimality. In the next section we will show how to parameterize the set of smooth orbit equivalence classes of area-preserving flows in such a way that in the absence of homologically trivial closed orbits most flows have no saddle connection. Right now we will show that in natural one-parameter families of area-preserving flows similar to linear flows on the octagon with slope as a parameter, all but countably many flows have no saddle connection.

First, let us generalize the octagon construction from Section 14.4b. Let $P \subset \mathbb{R}^2$ be a polygon whose angles may be less than, equal to, or greater than π. In other words, P is an ordinary (not necessarily convex) polygon and some of its sides may be artificially subdivided into several pieces. Furthermore assume that the sides of P are divided into pairs of parallel arcs of equal length.

Denote the translations identifying the sides in each pair by $T_1^{\pm 1}, \ldots, T_m^{\pm 1}$. Any centrally symmetric polygon, convex or not, with pairs of opposite sides identified is an example of such an arrangement. Another example is given by the L-shaped polygon with two vertices added on the longer (outer) sides. Identifying the pairs of sides by the translations makes P into a compact closed surface $\tilde{P}$. As in Section 14.4b there is an obvious smooth structure at all points except for those that come from vertices. At these points the smooth structure can also be defined, but it is not essential for describing the orbit structure of the parallel flow. For each value of the angle $\alpha \in [0, 2\pi)$ one can define the parallel flow on $\tilde{P}$ as the motion with unit speed along the oriented lines that form an angle α with the fixed direction (and taking into account the identification). Such a flow is discontinuous; it is defined for all values of time only at points whose orbits never hit a vertex. Since the set of such points has full Lebesgue measure λ it is defined as a measure-preserving flow for all values of time. In order to make Theorem 14.6.3 applicable, multiply the vector field X_α defining the flow by a nonnegative function ρ vanishing precisely at the vertices and such that ρ^{-1} is Lebesgue integrable. The vector field ρX_α is C^0 and integrates uniquely to a C^0 flow preserving the measure $\rho^{-1}\lambda$ which is positive on open sets. The vertices are saddle-type fixed points and the first-return map on any transversal coincides with that for the original discontinuous flow. Denote by $\mathcal{T}$ the group of parallel translations generated by the translations $T_1, \ldots, T_m$. Let $p_1, \ldots, p_{2m}$ be the vertices of P.

Proposition 14.6.5. *If a parallel flow on $\tilde{P}$ has a saddle connection then its direction is parallel to a vector of the form $g + p_j - p_i$ for some $g \in \mathcal{T}$, $i, j = 1, \ldots, 2m$.*

Proof. Suppose there is a saddle connection, that is, an orbit that starts at p_i and ends at p_j. Consider the following process of "unfolding". Start at p_i and each time the orbit reaches a side, instead of applying the appropriate translation to the point, apply its inverse to the entire polygon. This way we obtain a correspondence between the orbit and a segment of the straight line beginning at p_i. Since there is a saddle connection, after finitely many crossings the segment will reach a vertex of a shifted polygon $T\tilde{P}$. Obviously the shift T is a linear combination of basic translations T_k, $k = 1, \ldots, m$, with integer coefficients and the vertex is the translate of p_j. Thus the orbits are parallel to $T + p_j - p_i$. □

Corollary 14.6.6. *For all but countably many values of α the linear flow on $\tilde{P}$ has no saddle connection and is hence quasi-minimal. The same applies to the interval exchange transformation induced by the flow on any straight line segment.*

At the end of Section 14.4 we described a correspondence between the family of parallel flows in the regular octagon and the billiard flow inside the right triangle with an angle $\pi/8$. This construction allows a generalization which we briefly sketch here.

Consider the billiard inside a polygon P whose angles are commensurable with π. We will call such polygons *rational*. Let G be the group of plane motions generated by the reflections in the sides of P. It has a finite-index normal subgroup G_0 of parallel translations and the factor group G/G_0 is isomorphic to a dihedral group D_r; it corresponds to the action of G on the set of directions. In other words, the direction of any billiard orbit after reflection belongs to the same orbit of G/G_0. Now one can pick elements $g_0, \dots, g_{2N-1}$ of G in each coset of G_0. They can be ordered in such a way that $g_0 = \mathrm{Id}$, $g_{m+1} = R_m g_m$, where R_m is one of the reflections in the sides of P generating G. Now we take the $2N$ copies $P, R_1P, R_2R_1P, \dots, R_{2N-1}\cdots R_1P$ of P and identify them using the corresponding reflections. The resulting figure may have overlaps, unlike the case of the 16 triangles in Figure 14.4.4. Nevertheless, for any "free" side of one of the polygons there is a free side of another one which can be identified by a translation from G_0. Thus one can construct a closed surface S and for any orbit of G/G_0 a flow on P lifts to a parallel flow on S which is defined unambiguously despite the presence of several "sheets". Let us consider the billiard-ball map on the set of unit tangent vectors with footpoint on ∂P. A convenient way to study it is to pick a side l and a direction α and consider the return map to $l \times (G/G_0)\{\alpha\}$. The latter is a union of $2N$ intervals and after putting them side by side and appropriately normalizing Lebesgue measure on each of them we obtain a family $I^{(\alpha)}$ of interval exchange transformations. Arguing exactly as in the proof of Proposition 14.6.5 and passing to the section we obtain

Proposition 14.6.7. *For any side l of a rational polygon the interval exchange transformation $I^{(\alpha)}$ has no saddle connection for all but countably many values of α.*[2]

We have seen that Poincaré maps on transversals to measure-preserving flows are isomorphic to interval exchange transformations. There is, in fact, a construction showing that at least in the oriented case any interval exchange transformation appears in this way, that is, for any oriented interval exchange transformation $I_{v,\pi}$ (see Definition 14.5.1) there exists a compact orientable surface M, a smooth area-preserving flow φ with finitely many fixed points of saddle type, and a transversal τ such that the first-return map of φ on τ is smoothly conjugate to $I_{v,\pi}$. Thus in this sense these theories are equivalent.

Exercises

14.6.1. *Construct an example of an area-preserving flow whose fixed points are simple saddles on the sphere with n handles that has n transitive components and no periodic components.*

14.6.2. *Consider the parallel flow in the L-shaped polygon with identifications as described near the beginning of Subsection b. Show that the resulting surface is homeomorphic to the sphere with 2 handles and after an appropriate time change the flow has one topological double saddle.*

14.6.3. *Consider the parallel flow in the L-shaped polygon with identifications as described above. Reduce this to a parallel flow in a polygon and calculate the genus of the resulting surface.*

7. Generalizations of rotation number

a. Rotation vectors for flows on the torus. In the discussion of circle diffeomorphisms in Chapter 12 we saw that the possibility of a smooth conjugacy to a linear map is related to the arithmetic properties of the rotation number of the circle diffeomorphism. As is to be expected, the situation is similar for flows on the torus. This is one of many motivations for developing a notion corresponding to rotation number. As in the definition of the rotation number of circle diffeomorphisms we need to pass to the universal cover of the space. However, there is an interesting distinction. For the circle the choice of generator in the first homology group is unique up to an orientation, but the lift of a map to the universal cover is defined up to a deck transformation and the latter factor is responsible for the rotation number being defined only modulo 1. For a vector field on the torus (or any manifold) the lift to the universal cover is unique, but the choice of generators in $H_1(\mathbb{T}^2, \mathbb{Z})$ is not. Accordingly we will be able to define a rotation *vector* for any given choice of basis in $H_1(\mathbb{T}^2, \mathbb{Z})$ and thus it is determined up to an action of $SL(2, \mathbb{Z})$.

Proposition 14.7.1. *Let φ^t be a fixed-point-free C^1 flow on $\mathbb{T}^2$ and denote the lift of φ^t to the universal cover $\mathbb{R}^2$ by Φ^t. Then for every $x \in \mathbb{R}^2$ the limit*

$$\rho(\varphi) := \lim_{t\to\infty} \frac{1}{t}\Phi^t(x) \in \mathbb{R}^2 \tag{14.7.1}$$

exists and is independent of x. We will call it the rotation vector of φ.

Proof. First notice that existence and independence of x of $\rho(\varphi)$ is an invariant of flow conjugacy. Let $h: \mathbb{T}^2 \to \mathbb{T}^2$ be a homeomorphism and $H = L + G$ its lift to the universal cover $\mathbb{R}^2$, where L is a linear map and G is periodic. Let $x \in \mathbb{R}^2$ and $y = H^{-1}(x)$. Then

$$\frac{1}{t}H(\Phi^t(H^{-1}(x))) = L\Big(\frac{1}{t}\Phi^t(y)\Big) + \frac{1}{t}G(\Phi^t(y)).$$

$G(\Phi^t(y))$ is bounded, so $\lim_{t\to\infty} H(\Phi^t(H^{-1}(x)))/t = L\lim_{t\to\infty}\Phi^t(y)/t$.

Using Proposition 14.2.1 we construct a closed transversal τ to the flow φ^t. By Proposition 14.2.2 we obtain a C^1 diffeomorphism $h: \mathbb{T}^2 \to \mathbb{T}^2$ which maps τ into the standard "horizontal" circle $\tau_0 := S^1 \times \{0\} = \{(s, 0) \mid s \in \mathbb{R}/\mathbb{Z}\}$ (we use additive notation). We will show existence of the limit (14.7.1) for the flow $h \circ \varphi^t \circ h^{-1}$. Since every point returns to τ_0 and the return time to τ_0 is bounded it is sufficient to show existence of the limit for points on τ_0. Furthermore by the same reason it is sufficient to consider only the sequence of moments $t_n(s)$ of returns to τ_0. Denote the return map to τ_0 by f and its lift to $\mathbb{R} \times \{0\}$ by F. Notice that on the universal cover a return to τ_0 corresponds to a change of the second coordinate by 1 or -1; without loss of generality we consider the first case. Then $\Phi^{t_n(s)}(s,0) = (F^n(s), n)$. Notice that $t_n(s) = t(s) + t(f(s)) + \cdots + t(f^{n-1}(s))$, where $t(s)$ is the return time to τ_0. We use existence of the rotation number for s, that is, $\lim_{n\to\infty} F^n(s)/n = \tau(f)$, and unique ergodicity of F, which implies $\lim_{n\to\infty} t_n(s)/n = \int t(s)\, d\mu =: t_0$, where μ is the unique invariant Borel probability measure for F. Then

$$\lim_{n\to\infty} \frac{\Phi^{t_n(s)}(s,0)}{t_n(s)} = \lim_{n\to\infty} \Big(\frac{n}{t_n(s)} \cdot \frac{F^n(s)}{n}, \frac{n}{t_n(s)}\Big) = \Big(\frac{\tau(f)}{t_0}, \frac{1}{t_0}\Big).$$

□

Now we can reformulate Corollary 14.2.7 without referring to a section.

Corollary 14.7.2. *Let φ^t be a C^∞ flow preserving an area element. If the coordinates of the rotation vector $\rho(\varphi)$ are Diophantine numbers then φ^t is C^∞ conjugate to a linear flow.*[1]

b. Asymptotic cycles.[2] Now consider a more general situation. Let M be a compact differentiable manifold and φ^t a C^1 flow generated by the vector field X and preserving a measure μ. Let ω be a closed differential 1-form. Then the integral

$$\int X \lrcorner \omega \, d\mu$$

in fact depends only on the cohomology class of ω, for if $\omega_2 - \omega_1 = dF$ then $X \lrcorner(\omega_2 - \omega_1) = \mathcal{L}_X F$ and

$$\int X \lrcorner \omega_2 \, d\mu - \int X \lrcorner \omega_1 \, d\mu = \int \mathcal{L}_X F \, d\mu = \Big(\frac{d}{dt} \int F \circ \varphi^t \, d\mu\Big)\Big|_{t=0} = 0$$

due to the preservation of μ. Thus the map $\omega \mapsto \int X \lrcorner \omega \, d\mu$ defines a linear functional on the first de Rham cohomology group of M which by duality can be identified with an element $\rho_\mu \in H_1(M, \mathbb{R})$, which is called the *asymptotic cycle* of the flow with respect to the measure μ.

Suppose now that μ is ergodic. In this case we can give a geometric interpretation of the asymptotic cycle. By the Birkhoff Ergodic Theorem 4.1.2

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \omega(X(\varphi^s(x)))\, ds = \int X \lrcorner \omega \, d\mu \tag{14.7.2}$$

for μ-a.e. $x \in M$. Let $\gamma_t(x)$ be the oriented orbit segment from x to $\varphi^t(x)$. By definition of integration of differential forms $\int_0^t \omega(X(\varphi^s(x)))\, ds = \int_{\gamma_t(x)} \omega$. Now we proceed similarly to the construction of homotopical entropy for a flow at the end of Section 3.1. Namely, pick a family of arcs $\gamma_{x,y}$ of bounded length

Replace $\gamma_t(x)$ by the closed loop $\tilde{\gamma}_t(x) := \gamma_t(x) \cdot \gamma_{y,x}$. From (14.7.2) and (14.7.3) we obtain

$$\lim_{t\to\infty} \frac{1}{t} \int_{\tilde{\gamma}_t(x)} \omega = \int X \lrcorner \omega \, d\mu.$$

Since any homology class in $H_1(M, \mathbb{R})$ is uniquely determined by its values on a basis of closed 1-forms we can deduce that for μ-a.e. $x \in M$

$$\rho_\mu = \lim_{t\to\infty} \frac{1}{t} [\tilde{\gamma}_t(x)],$$

where $[\cdot]$ denotes the homology class.

Note that for a uniquely ergodic flow all the above a.e. convergences are uniform. It is not difficult to see (Exercise 14.7.1) that the rotation vector for a fixed-point-free flow on the two-torus is simply a coordinate representation of the asymptotic cycle with respect to the standard basis in the first cohomology group.

In the two-dimensional orientable case which is currently our prime concern, we can also interpret the asymptotic cycles as elements of the first *cohomology group*. In general, for flows on n-dimensional oriented manifolds the corresponding elements belong to the $(n-1)$st cohomology group. Namely, a vector field X and an invariant measure μ define a flux *current*, an object similar to closed $(n-1)$-forms, which can be integrated over $(n-1)$-submanifolds. If τ is an $(n-1)$-dimensional oriented transversal to X and $A \subset \tau$ a Borel subset then the flux $\mathcal{F}(A)$ is defined as $\pm\mu(\bigcup_{t=0}^{\epsilon} \varphi^t(A))/\epsilon$, where φ^t is the flow generated by X, $\epsilon > 0$ is any small number, and the sign is determined according to the agreement of the orientation of M and the orientation obtained from the orientation of X and that on τ. To simplify the discussion we consider the special case when the measure μ is given by the volume Ω and that the vector field is C^1. Then the flux current is defined by integrating the $(n-1)$-form $X \lrcorner \Omega$ called the *flux form*.

Lemma 14.7.3. *The flux form is closed if and only if the vector field X preserves Ω.*

Proof. By (A.3.3) and $d\Omega = 0$ we have $\pounds_X \Omega = d(X \lrcorner \Omega)$. □

Any closed $(n-1)$-form ω determines a linear functional l_ω in $H^1(M, \mathbb{R})$ (that is, on $H_1(M, \mathbb{R})$) via $l_\omega(\alpha) = \int_M \omega \lrcorner \alpha$. Applying this to our case $\omega = X \lrcorner \Omega$ we obtain

$$l_{X\lrcorner\Omega}(\alpha) = \int_M (X \lrcorner \Omega) \wedge \alpha = \int_M (X \lrcorner \alpha) \Omega = \int_M (X \lrcorner \alpha) \, d\mu = \rho_\mu(\alpha).$$

Next we will show how in the two-dimensional area-preserving case the asymptotic cycle can be extended to an invariant giving a complete local (in the space of vector fields) classification up to smooth orbit equivalence.

c. Fundamental class and smooth classification of area-preserving flows.[3] We say that a zero p of an area-preserving vector field on a surface is a *generic saddle of index* $-n$ (or a *generic* n*-fold saddle*) if in local coordinates near p the vector field is Hamiltonian with Hamiltonian function $H(x,y) = \prod_{i=1}^{n+1}(\alpha_i x - \beta_i y) + R(x,y)$, where all ratios β_i/α_i are different and R has zero $(n+1)$-jet at 0. Thus any standard linear saddle is a generic 1-fold saddle and the saddle shown at the left of Figure 8.4.1 is a generic threefold saddle.

Consider a closed compact orientable surface M of genus g, a smooth 2-form Ω on M, $p_1, \dots, p_r \in M$, and $n_1, \dots, n_r \in \mathbb{N}$ such that $\sum_{i=1}^r n_i = 2g-2$. Let $\mathcal{X}$ be the space of C^∞ vector fields on M preserving the form Ω such that the point p_i is a generic saddle of index $-n_i$ for $i \in \{1, \dots, r\}$ and there are no other zeros. This description of critical points agrees with the Poincaré–Hopf Index Formula (Theorem 8.6.6). We call $\Delta := \{p_1, \dots, p_r\}$ the *critical set* and consider the space of 1-cycles on M with real coefficients *relative to* Δ. Such a cycle can be represented as a linear combination of oriented arcs in M whose boundaries belong to Δ. Relative boundaries are the same as ordinary boundaries. Thus the space of relative cycles factored by relative boundaries has dimension $2g+r-1$; one needs to add to $2g$ independent cycles a collection of arcs connecting the points in Δ and forming a tree.

The restriction of the flux form for any vector field $X \in \mathcal{X}$ to the space of relative cycles is called the *fundamental class* of X and will be denoted by $FC(X)$. The first classification result for area-preserving vector fields on orientable surfaces of genus $g \geq 2$ can be summarized as follows.

Theorem 14.7.4. *Suppose $X_t \in \mathcal{X}$, $0 \leq t \leq 1$, is a smooth family such that $FC(X_t) = \lambda_t FC(X_0)$, where λ_t is a positive scalar. Then there exists a family $h_t\colon M \to M$ of Lipschitz homeomorphisms that are C^∞ diffeomorphisms away from the critical set, and a positive function μ_t such that $h_t(p_i) = p_i$ for $i = 1, \dots, r$ and $(h_t)_* X_0 = \mu_t X_t$. In other words, h_t effects a Lipschitz orbit equivalence between the flows generated by X_0 and X_t that is C^∞ away from the critical set.*

Proof. Not surprisingly we will use a version of the "homotopy trick" that was first used in the proof of the Moser Theorem 5.1.27 and then appeared several more times. Consider the one-form $\omega_t = X_t \lrcorner \Omega$ and let $\alpha_t := \dfrac{d\omega_t}{dt}$. We will look for the infinitesimal generator $H_t := \dfrac{dh_t}{dt}$ of the family h_t. Suppose we found h_t such that $h_t^* \omega_t = \omega_0$. Then if $h_t^*\Omega = \lambda_t \Omega$ we have

$$X_0 \lrcorner \Omega = h_t^*(X_t \lrcorner \Omega) = (h_t)_* X_t \lrcorner h_t^* \Omega = \lambda_t^{-1}(h_t)_* X_t \lrcorner \Omega,$$

that is, $(h_t)_* X_t = \lambda_t X_0$. By (A.3.3) we have

$$\frac{d}{dt}(h_t^* \omega_t) = h_t^* \mathcal{L}_{H_t} \omega_t + h_t^* \alpha_t = h_t^*(d(H_t \lrcorner \omega_t) + \alpha_t),$$

so we want to find a vector field H_t for which the right-hand side of this identity vanishes.

By assumption α_t is an exact form, that is, $\alpha_t = dP_t$, where the function P_t is defined up to a constant. Furthermore, since α_t vanishes on relative cycles $P_t(p_i) = P_t(p_j)$ for $i, j = 1, \dots, r$, we can assume that $P_t(p_i) = 0$ for all i. Thus it suffices to solve

$$H_t \lrcorner \omega_t = -P_t. \tag{14.7.4}$$

It follows from the definition that $\ker \omega_t = X_t$. The solution of (14.7.4) is defined up to a term from $\ker \omega_t$. Fixing a Riemannian metric we define the solution uniquely by making it orthogonal to X_t and such that (X_t, H_t) is a positively oriented pair. Naturally $H_t(p_i) = 0$. Thus H_t is defined and continuous and C^∞ away from the zeros of X_t. At the point p_i the form w_t has zeros of order n_i and the same is true for α_t since the Taylor coefficients are differentiable in t. Hence P_t has zeros of order $n_i + 1$ and H_t chosen along the gradient lines of p_i decreases near p_i in proportion to the distance to p_i. Hence the H_t are Lipschitz vector fields and H_t is uniquely integrable to a one-parameter family of Lipschitz homeomorphisms that are smooth away from the critical set and define orbit equivalences between X_t and X_0. □

Remark. The source of nonsmoothness at the critical set is the presence of local invariants of smooth orbit equivalence near multiple saddles. Namely, an n-fold saddle has $2n+2$ separatrices and if two such saddles are smoothly orbit equivalent then the tangent directions at the separatrices must be carried to each other by the derivative of the conjugacy. Thus we have to consider the action of $GL(n, \mathbb{R})$ on $(n+1)$-tuples of lines. This action is transitive for $n \leq 2$, but there are invariants (cross-ratios) for $n \geq 3$.

In fact, if all saddles are no more than double then in Theorem 14.7.4 smooth orbit equivalence can be achieved. First notice that if the vector fields are identical near the critical set then the resulting conjugacy will be the identity nearby as well. Thus one can first find a local coordinate change near each critical point that brings the saddle into a standard form. For the case of a simple saddle this is a continuous-time counterpart of Exercises 6.6.4–6.6.5. Making a time change and carefully applying the Moser Theorem 5.1.27 we reduce to the situation of flows that are identical near the critical set.

Proposition 14.7.5. *Consider an area-preserving vector field on a surface with finitely many fixed points of the saddle type. Then invariant measures supported on transitive components are determined uniquely by their asymptotic cycles.*

Remark. This is an analog of Theorem 14.5.14 for flows.

Proof. By Theorem 14.6.3 it suffices to show that the flux through a small transversal inside a transitive component is determined by the asymptotic cycle of the measure, that is, by the fluxes through closed curves. To that end we take, using Theorem 14.6.3, an orbit segment starting very close to one end of the transversal and ending very close to the other end, and close it using the piece of transversal. The fluxes through these closed curves coincide with those through their transverse portion and on the other hand each of them and hence their limit, which is also the flux through the transversal, is indeed determined by the asymptotic cycle. □

Theorem 14.7.6. *There are at most* genus(M) *nontrivial ergodic invariant measures for any area-preserving vector field on a surface M.*[4]

Proof. We begin by showing that the asymptotic intersection number (see the remarks after Theorem A.5.2) of any two dense orbits for such a vector field is zero. To that end take a transversal and for any two orbits consider segments beginning and ending on the transversal, and intersecting the transversal n times. We refer to the construction of the asymptotic cycle by closing orbits by transverse pieces and close these two orbit segments by the pieces of the transversal connecting their ends. The lengths of the resulting curves then are of order n each. They can intersect each other no more than $2n$ times, namely, on the transversal, so their intersection number is of order $2n/n^2$ after normalizing by length, whence the limit is indeed zero. Thus all dense orbits lie in a g-dimensional Lagrangian subspace for the (symplectic) intersection form (see the remarks after Theorem A.5.2).

Now nontrivial flow-invariant measures are determined by their asymptotic cycle, that is, we have an injective map from ergodic measures to asymptotic cycles. It is affine and hence mutually singular measures correspond to linearly independent points as in the proof of Theorem 14.7.6; hence the image lies in this g-dimensional subspace, so there are at most g distinct ergodic measures. □

Exercises

14.7.1. *Show that the rotation vector for a fixed-point-free flow on the two-torus is the coordinate representation of the asymptotic cycle with respect to the standard basis in the first cohomology group.*

14.7.2*. *Consider the double torus and pick a standard basis $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ in the first homology group that consists of generators of the homology of one of the joined tori. Pick two points p, q and join them by a short curve γ_5. Show that for any (x_1, x_2, x_3, x_4) with positive coordinates there exists an $\epsilon > 0$ such that for $|x_5| < \epsilon$ there exists a C^∞ area-preserving flow on the double torus with simple saddles at p and q, and the flux through γ_i is x_i.*

14.7.3. *Under the assumptions of the previous exercise show that if $(x_1, \ldots, x_5)$ are rationally independent then the resulting flow is quasi-minimal.*

14.7.4*. *Show that among the flows constructed in the previous exercises there are quasi-minimal nonergodic ones.*