CHAPTER 3

Hyperbolic Systems–general facts

This chapter is design to give an idea of the general results of hyperbolic theory. Since such a theory covers a rather vast landscape and it contains very technical results our exposition is bound to be quite sketchy. Nevertheless, in the following chapters we will prove all the results we need in this book for the special setting of area preserving two dimensional maps (see section 4.4 and Problem3.6 for Oseledets and chapter 7 for the results on foliations).

3.1 Hyperbolicity

Our goal in this section is to introduce and discuss a class of systems for which we can hope to extend the results of the previous section. As we have seen, the chief property the we used in the study of the Arnold cat were the expanding and contracting properties of the map. These are generalized in the following definition.

Definition 3.1.1 *By Hyperbolic System (with discrete time) we mean a Dynamical System (X, T, μ) such that X is a smooth compact Riemannian manifold (possibly with boundary), T is almost everywhere differentiable and there exists two measurable families of invariant*[1] *subspaces $E^u(x)$, $E^s(x) \in \mathcal{T}_x X$ transversal at almost each point,*[2] *and measurable functions $\nu(x) > 1$, $c(x) > 0$ such that for almost all $x \in X$*

$$\|D_x T^n v\| \geq c(x)^{-1}\nu(x)^n \|v\| \quad \forall v \in E^u(x)$$
$$\|D_x T^n v\| \leq c(x)\nu(x)^{-n} \|v\| \quad \forall v \in E^s(x).$$

If the functions c, ν can be chosen constant and the distributions are transversal at each point, then the system is called Uniformly Hyperbolic. In addition, T is a diffeomorphism and E^u, E^s vary with continuity, then the system is called Anosov (or sometimes C or U systems).

The condition in Definition 3.1.1 is essentially equivalent to saying that two very close initial conditions almost certainly will grow apart at an exponential rate. This corresponds to a strong instability with respect to the initial conditions and characterizes the sense in which the dynamics of hyperbolic systems is a very complex one. Such complex behaviour has recently captured the popular fantasy under the ambiguous name of chaos.

[1] That is $D_x T E^{s(u)}(x) = E^{s(u)}(Tx)$.
[2] That is, $E^u(x) \cap E^s(x) = \{0\}$ and $E^u(x) \oplus E^s(x) = \mathcal{T}_x X$ a.e.

3.1.1 Examples

Rotations

Clearly the rotations are not hyperbolic since $DT = 1$.

Dilation

One can easily see that such a system is expanding, hence $E^u = \mathbb{R}$ and $E^s = \emptyset$.

Arnold cat

We have seen it in detail in the previous chapter.

Baker

In this case one direction is expanding and one is contracting, $\dim E^u = \dim E^s = 1$

A more general notion of hyperbolicity is the one of *hyperbolic set*.

Definition 3.1.2 *Given a diffeomorphism T of a manifold X, we say that $\Lambda \subset X$ is hyperbolic if Λ is compact, $T(\Lambda) = \Lambda$ and there exists two measurable families of invariant subspaces $E^u(x)$, $E^s(x) \in \mathcal{T}_x X$ transversal at each point and measurable functions $\nu(x) > 1$, $c(x) > 0$ such that for all $x \in \Lambda$*

$$\begin{aligned} \|D_x T^n v\| &\geq c(x)^{-1}\nu(x)^n \|v\| \quad \forall v \in E^u(x) \\ \|D_x T^n v\| &\leq c(x)\nu(x)^{-n} \|v\| \quad \forall v \in E^s(x). \end{aligned}$$

If the constants c, ν can be chosen independently of $x \in \Lambda$ then Λ is called Uniformly Hyperbolic.

3.1.2 Examples

Solenoid

...

Smale Horseshoe

In this case the set Λ is the one constructed in Examples 1.4.1 and $\dim E^s = \dim E^u = 1$.

Forced pendulum

Same situations as for the horseshoe, see Examples 1.8.1.

Definition 3.1.1 it is not particularly helpful in concrete cases since, in general, it is not clear how to verify if a systems is hyperbolic or not. A first step toward a better understanding is contained in the next section, then in chapter 4 we will study this issue further.

3.2 Lyapunov exponents and invariant distributions

We start by a different and very helpful characterization of hyperbolicity obtained by introducing the so called Lyapunov Exponents (LE).

Definition 3.2.1 *For each $x \in X$, $v \in \mathcal{T}_x X$ we define*

$$\lambda(x,\, v) = \lim_{n\to\infty} \frac{1}{n} \log \|D_x T^n v\|.$$

If $\lambda(x,\, v)$ exists it is called "Lyapunov exponent" (LE).

It is interesting to notice that $\lambda(Tx, D_xTv) = \lambda(x,v)$ (see Problem 3.1). Moreover it should be clear that, if the system is ergodic and the map invertible, then $\lambda(x,v)$, if it exists, can assume only finitely many values (see Problem 3.3).

The existence and properties of the LE have been studied in detail. Here we satisfy ourselves with a basic fact.

Theorem 3.2.2 (Oseledec [57]) *For each (X, T, μ), X finte union of compact Riemannian manifolds, T almost everywhere differentiable, if*

$$\int_X \| \log D_x T \| d\mu < \infty, \tag{3.2.1}$$

then, for almost all $x \in X$, there exists numbers $\{\lambda_1, \ldots \lambda_{s(x)}\}$ and a flag of subspaces

$$\{0\} = V_0 \subset V_1(x) \subset \ldots \subset V_{s(x)-1}(x) \subset V_{s(x)}(x) = \mathcal{T}_x X,$$

such that, for all $v \in V_k(x) \backslash V_{k-1}(x)$, the LE $\lambda(x,v)$ exists and equal λ_k.

Note that if T is invertible, $\{\lambda_i(x)\}$ is equal a.e. to $\{-\lambda_i^-(x)\}$ where $\{\lambda_i^-(x)\}$ are the LE of (X, T^{-1}, μ) (see Problem 3.4).

We will not prove Theorem 3.2.2 in the above generality, but see section 4.4 for a very constructive proof in an important, but special, case.

To appreciate the relevance of the LE, in the present context, consider the following.

Theorem 3.2.3 *A system $(X,\, T,\, \mu,)$, where X is a Riemannian manifold and T is a diffeomorphism, is hyperbolic iff for almost all $x \in X$*

$$\lambda(x,\, v) \neq 0 \quad \forall v \in \mathcal{T}_x X, \;\; v \neq 0.$$

Proof. Clearly, if the system is hyperbolic, then all the LE are different from zero. The other implication is almost as trivial. Define $E^s(x) = \{v \in \mathcal{T}_x \mid \lambda(x,v) < 0\}$; then consider the Dynamical system (X, T^{-1}, μ) and its LE $\lambda^-(x,v)$ and define $E^u(x) = \{v \in \mathcal{T}_x \mid \lambda^- < 0\}$. Next, let

$$\rho(x) = \sup\{\lambda(x,v),\, \lambda_-(x,w) \mid v \in E^s(x),\, w \in E^u(x)\}$$

clearly $\rho(x) < 0$ a.e.. Then setting $\nu(x) = e^{-\rho(x)/2}$ and

$$c(x) = \sup\{\nu(x)^n \|D_x T^n v\|,\, \nu(x)^n \|D_x T^{-n} w\| \mid v \in E^s(x);\; w \in E^u(x)\}_{n\in\mathbb{N}},$$

the theorem is proven. □

3.3 Invariant manifold of a fixed point

In order to gain a feeling for the type of results we will present in the next section let us consider the simplest possible case in which the existence of invariant manifolds arises: the Hadamard-Perron theorem.

Definition 3.3.1 *Given a smooth map $T : X \to X$, X being a Riemannian manifold, and a fixed point $p \in X$ (i.e. $Tp = p$) we call (local) stable manifold (of size δ) a manifold $W^s(p)$ such that*[3]

$$W^s(p) = \{x \in B_\delta(x) \subset X \mid \lim_{n\to\infty} d(T^n x, p) = 0\}.$$

Analogously, we will call (local) unstable manifold (of size δ) a manifold $W^u(p)$ such that

$$W^u(p) = \{x \in B_\delta(x) \subset X \mid \lim_{n\to\infty} d(T^{-n} x, p) = 0\}.$$

It is quite clear that $TW^s(p) \subset W^s(p)$ and $TW^u(p) \supset W^u(p)$ (Problem 3.7). Less clear is that these sets deserve the name "manifold." Yet, if one thinks of the Arnold cat at the point zero (which is a fixed point) it is obvious that the stable and unstable manifolds at zero are just segments in the stable and unstable direction, the next Theorem shows that this is a quite general situation.

Theorem 3.3.2 (Hadamard-Perron) *Consider an invertible map $T : U \subset \mathbb{R}^2 \to U$, twice differentiable, such that $T0 = 0$ and*

$$D_0T = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \tag{3.3.2}$$

where $0 < \mu < 1 < \lambda$.[4] *That is, the map T is hyperbolic at the fixed point 0. Then there exists stable and unstable manifolds at 0. Moreover, $\mathcal{T}_0 W^{s(u)}(0) = E^{s(u)}(0)$ where $E^{s(u)}(0)$ are the expanding and contracting subspaces of D_0T.*

Proof. We will deal explicitly only with the unstable manifold since the stable one can be treated exactly in the same way by considering T^{-1} instead of T.

Since the map is twice differentiable we use the Taylor formula to obtain a convenient representation is T in a 2δ-neighborhood of zero and write

$$T(x) = D_0Tx + R(x) \tag{3.3.3}$$

where $\|R(x)\| \leq \frac{C}{2}\|x\|^2$, $\|D_x R\| \leq C\|x\|$, and $C = \sup_{\|x\|\leq 2\delta} \|D_x^2 R\|$.

Existence–a fixed point argument

The first step is to decide how to represent manifolds. In the present case, since we deal only with curves, it seems very reasonable to consider the set of curves $\Gamma_{\delta,c}$ passing through zero

[3] Sometime we will write $W^s_\delta(p)$ when the size really matters. By $B_\delta(x)$ we will always mean the open ball of radius δ centered at x.

[4] Notice that if D_0T has eigenvalues $0 < \mu < 1 < \lambda$ then one can always perform a change of variables such that (3.3.2) holds.

and "close" to being horizontal, that is the differentiable functions $\gamma : [-\delta, \delta] \to \mathbb{R}^2$ of the form

$$\gamma(t) = \begin{pmatrix} t \\ u(t) \end{pmatrix}$$

and such that $\gamma(0) = 0$; $\|(1,0) - \gamma'\|_\infty \leq c\delta$. It is immediately clear that any smooth curve passing through zero and with tangent vector, at each point, in the cone $\mathcal{C} := \{(a, b) \in \mathbb{R}^2 \mid |\frac{b}{a}| \leq c\delta\}$, can be associated to a unique element of $\Gamma_{\delta,c}$, just consider the part of the curve contained in the strip $\{(x, y) \in \mathbb{R}^2 \mid |x| \leq \delta\}$. Moreover, if $\gamma \in \Gamma_{\delta,c}$ then $\gamma \subset B_{2\delta}(0)$, provided $c \leq \delta^{-1}$.

Notice that it suffice to specify the function u in order to identify uniquely an element in $\Gamma_{\delta,c}$. It is then natural to study the evolution of a curve through the change in the associated function.

To this end let us investigate how the image of a curve in $\Gamma_{\delta,c}$ under T looks like.

$$T\gamma(t) = \begin{pmatrix} \lambda t + R_1(t, u(t)) \\ \mu u(t) + R_2(t, u(t)) \end{pmatrix} := \begin{pmatrix} \alpha_u(t) \\ \beta_u(t) \end{pmatrix}.$$

At this point the problem is clearly that the image it is not expressed in the way we have chosen to represent curves, yet this is easily fixed. First of all, $\alpha_u(0) = \beta_u(0) = 0$. Second, by choosing $\delta < \frac{\lambda}{4C}$, we have $\alpha_u'(t) > 0$, that is, α_u is invertible. In addition, $\alpha_u([-\delta, \delta]) \supset [-\lambda\delta + C\delta^2, \lambda\delta - C\delta^2] \supset [-\delta, \delta]$, provided $\delta \leq \frac{\lambda-1}{C}$. Hence, α_u^{-1} is a well defined function from $[-\delta, \delta]$ to itself. Finally,

$$|\frac{d}{dt}\beta_u \circ \alpha_u^{-1}(t)| = \left|\frac{\beta_u'(\alpha_u^{-1}(t))}{\alpha_u'(\alpha_u^{-1}(t))}\right| \leq \frac{\mu c\delta + 4C\delta}{\lambda - 4C\delta} \leq c\delta$$

where, again, we have chosen $\delta \leq \frac{\lambda-\mu}{8C}$ and $c \geq \frac{8C}{\lambda-\mu}$.

We can then consider the map $\tilde{T} : \Gamma_{\delta,c} \to \Gamma_{\delta,c}$ defined by

$$\tilde{T}\gamma(t) := \begin{pmatrix} t \\ \beta_u \circ \alpha_u^{-1}(t) \end{pmatrix} \tag{3.3.4}$$

which associates to a curve in $\Gamma_{\delta,c}$ its image under T written in the chosen representation. It is now natural to consider the set of functions $B_{\delta,c} = \{u \in \mathcal{C}^{(1)}([-\delta, \delta]) \mid u(0) = 0, |u'|_\infty \leq c\delta\}$ in the vector space $\mathrm{L}ip([-\delta, \delta])$.[5] As we already noticed $B_{\delta,c}$ is in one-one correspondence with $\Gamma_{\delta,c}$, we can thus consider the operator $\hat{T} : \mathrm{L}ip([-\delta, \delta]) \to \mathrm{L}ip([-\delta, \delta])$ defined by

$$\hat{T}u = \beta_u \circ \alpha_u^{-1} \tag{3.3.5}$$

From the above analysis follows that $\hat{T}(B_{\delta,c}) \subset B_{\delta,c}$ and that $\hat{T}u$ determines uniquely the image curve.

The problem is then reduced to studying the map $\hat{T}$. The easiest, although probably not the most productive, point of view is to show that $\hat{T}$ is a contraction in the sup norm. Note that this creates a little problem since $\mathcal{C}^{(1)}$ it is not closed in the sup norm (and not even $\mathrm{L}ip([-\delta, \delta])$ is closed). Yet, the set $B^*_{\delta,c} = \{u \in \mathrm{L}ip([-\delta, \delta]) \mid u(0) = 0, \sup_{t,s\in[-\delta,\delta]} \frac{|u(s)-u(t)|}{|t-s|} < c\}$ is closed (see Problem 3.8). Thus $\overline{B_{\delta,c}} \subset B^*_{\delta,c}$. This means

[5] This are the Lipschitz functions on $[-\delta, \delta]$, that is the functions such that $\sup_{t,s\in[-\delta,\delta]} \frac{|u(s)-u(t)|}{|t-s|} < \infty$.

that, if we can prove that the sup norm is contracting, then the fixed point will belong to $B^*_{\delta,c}$ and we will obtain only a Lipschitz curve. We will need a separate argument to prove that the curve is indeed smooth.

Let us start to verify the contraction property. Notice that

$$\alpha_u^{-1}(t) = \lambda^{-1}t + \lambda^{-1}R_1(\alpha_u^{-1}(t), u(\alpha_u^{-1}(t))),$$

thus, given $u_1, u_2 \in B_{\delta,c}$, by Lagrange Theorem

$$\begin{aligned}|\alpha_{u_1}^{-1}(t) - \alpha_{u_2}^{-1}(t)| &\leq \lambda^{-1}|\langle\nabla_\zeta R_1,\, (\alpha_{u_1}^{-1}(t) - \alpha_{u_2}^{-1}(t), u_1(\alpha_{u_1}^{-1}(t)) - u_2(\alpha_{u_2}^{-1}(t)))\rangle| \\ &\leq \frac{2C\delta}{\lambda}\left\{2|\alpha_{u_1}^{-1}(t) - \alpha_{u_2}^{-1}(t)| + |u_1(\alpha_{u_2}^{-1}(t)) - u_2(\alpha_{u_2}^{-1}(t))|\right\}.\end{aligned}$$

This implies immediately

$$|\alpha_{u_1}^{-1}(t) - \alpha_{u_2}^{-1}(t)| \leq \frac{4\lambda^{-1}C\delta}{1 - 2\lambda^{-1}C\delta}\|u_1 - u_2\|_\infty. \tag{3.3.6}$$

On the other hand

$$\begin{aligned}|\beta_{u_1}(t) - \beta_{u_2}(t)| &\leq \mu|u_1(t) - u_2(t)| + |\langle\nabla_\zeta R_2,\, (0, u_1(t) - u_2(t))\rangle| \\ &\leq (\mu + C\delta)\|u_1 - u_2\|_\infty. \end{aligned}\tag{3.3.7}$$

Moreover,

$$|\beta_u'(t)| \leq \mu + 2C\delta. \tag{3.3.8}$$

Collecting the estimates (3.3.6, 3.3.7, 3.3.8) readily yields

$$\begin{aligned}\|\hat{T}u_1 - \hat{T}u_2\|_\infty &\leq \|\beta_{u_1}\circ\alpha_{u_1}^{-1} - \beta_{u_1}\circ\alpha_{u_2}^{-1}\|_\infty + \|\beta_{u_1}\circ\alpha_{u_2}^{-1} - \beta_{u_2}\circ\alpha_{u_2}^{-1}\|_\infty \\ &\leq \left\{[\mu + 2C\delta]\frac{4\lambda^{-1}C\delta}{1 - 2\lambda^{-1}C\delta} + (\mu + C\delta)\right\}\|u_1 - u_2\|_\infty \\ &\leq \sigma\|u_1 - u_2\|_\infty,\end{aligned}$$

for some $\sigma \in (0,1)$, provided δ is chosen small enough.

Clearly, the above inequality immediately implies that there exists a unique element $\gamma_* \in \Gamma_{\gamma,c}$ such that $\tilde{T}\gamma_* = \gamma_*$, this is the *local* unstable manifold of 0.

Regularity–a cone field

As already mentioned, a separate argument it is needed to prove that γ_* is indeed a $\mathcal{C}^{(1)}$ curve.

To prove this, one possibility would be to redo the previous fixed point argument trying to prove contraction in $\mathcal{C}^{(1)}_{Lip}$ (the $\mathcal{C}^{(1)}$ functions with Lipschitz derivative); yet this would require to increase the regularity requirements on T. A more geometrical, more instructive and more inspiring approach is the following.

Define the cone field $\mathcal{C}_{\theta,h}(x,u) := \{\xi \in B_h(x) \mid (a,b) = \xi - x;\ a \neq 0;\ |\frac{b}{a} - u| \leq \theta\}$, with $|u| \leq c\delta$, $\theta \leq c\delta$ and $h \leq \delta$. By construction $B_h(x) \cap \gamma_* \subset \mathcal{C}_{c\delta,h}$ for each $x \in \gamma_*$. We will study the evolution of such a cone field on γ_*.

For all $\xi \in \mathcal{C}_{\theta,h}(x,u)$, if $(a,b) = \xi - x$ and $(\alpha,\beta) = T\xi - Tx$, it holds

$$(\alpha,\beta) = D_xT(a,b) + \mathcal{O}(C\|(a,b)\|^2).$$

Thus, setting $(\alpha', \beta') = D_x T(a,b)$ and $u' = \frac{\beta'}{\alpha'}$,

$$\left|\frac{\beta}{\alpha} - u'\right| \leq \mu\lambda^{-1}[c_1 h + \theta],$$

for some constant c_1 depending only on T and δ. Accordingly, if h is small enough, there exists $\sigma \in (0,1)$ such that

$$T\mathcal{C}_{\theta,h}(x,u) \subset \mathcal{C}_{\sigma\theta,h}(Tx,u').$$

Hence, if $x \in \gamma_*$, $\gamma_* \cap B_{\sigma^{-n}h}(T^{-n}x) \subset \mathcal{C}_{c\delta,h}(T^{-nx},0)$ and, since $T^{-n}\gamma_* \subset \gamma_*$,

$$\gamma_* \cap B_h(x) \subset \mathcal{C}_{\sigma^n c,h}(x, v_n) \tag{3.3.9}$$

where $(a, av_n) = D_{T^{-n}x}T^n(1,0)$ and $h < \sigma^n c\delta$.

The estimate (3.3.9) clearly implies

$$\gamma_*'(x) = (1, \lim_{n\to\infty} v_n) \tag{3.3.10}$$

which indeed exists (see Problem 3.9). □

There is an issue not explicitly addresses in our formulation of Hadamard-Perron theorem: the uniqueness of the manifolds. It is not hard to prove that $W^{s(u)}(p)$ are indeed unique (see Problem 3,2).

The point of view employed in the previous theorem is brought to its extreme consequences in the Pugh-shub.

There is another point of view to study stable and unstable manifolds: to "grow" the manifolds. This is done by starting with a very short curve in $\Gamma_{\delta,c}$, e.g. $\gamma_0(t) = (t,0)$ for $t \in [\lambda^{-n}\delta,\, \lambda^n\delta]$, and showing that the sequence $\gamma_n := T^n\gamma_0$ converges to a curve in the strip $[-\delta,\, \delta]$, independent of γ_0. From a mathematical point of view, in the present case, it corresponds to spell out explicitly the proof of the fixed point theorem. Nevertheless, as we will see in ???, it is a more suggestive point of view and it is more convenient when the hyperbolicity is non uniform. For example consider the map[6]

$$T\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} 2x - \sin x + y \\ x - \sin x + y \end{pmatrix} \tag{3.3.11}$$

then 0 is a fixed point of the map but

$$D_0 T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is not hyperbolic, yet, due to the higher order terms, there exist stable and unstable manifolds (see Problems 3.11, 3.12, 3.13).

3.4 Invariant manifolds and foliations

The concept of stable and unstable manifold of a fixed point can be obviously generalized to periodic orbits and, less obviously but more interestingly, to <u>any</u> orbit:

[6] Some times this is called *Levowich map*.

Definition 3.4.1 *For each point $x \in X$ we define its stable (unstable) sets as*

$$W^{s(u)}(x) = \{z \in X \mid \lim_{n\to\infty(-\infty)} d(T^n x, T^n z) = 0\}.$$

Remark 3.4.2 *It will not surprise the reader that very often such sets are indeed manifolds, in this case they are called stable (unstable)* manifolds. *In the following we will discuss only this case.*

Very often the structure of the stable manifolds is quite complex (as we have already seen in the introduction) and for this reason it may be more convenient to talk about *local* stable (unstable) manifolds as we did already (see Definition 3.3.1).

Definition 3.4.3 *By* local stable (unstable) manifolds *at $x \in X$ we mean the connected component containing x of $B_\varepsilon(x) \cup W^{s(u)}(x)$, for some $\varepsilon > 0$.*

Clearly the local manifolds are neve unique, since their size may vary, but they may be unique once the size is assigned. Sometimes the noation $W^s_{\text{loc}}(x)$ as well as $W^s_\delta(x)$, we will use the former only if some confusion can arise and the latter only if the size plays an explicit rôle in the argument.

Definition 3.4.4 *By $W^s_\delta(x)$ we mean a local stable manifolds such that $\partial W^s_\delta(x) \subset \partial B_\delta(x)$.*

For simplicity we will restrict ourself to stable manifolds untill the end of the section, note that if the map is invertible, then the same considerations hld for the unstable manifold as well.

We will be mostly interested in the case in which almost every point has a stable manfold. Note that is $W^s(x)$ is a local stable manifold at x, then $TW^s(x)$ is a local stable manifold at Tx. This means that we have an invariant foliation

Definition 3.4.5 *By a foliation we mean*

The first interesting fact is the following.

Proposition 3.4.6 *If we have a stable foliation, then for almost all ponts $x \in X$ the associated stable manifold contracts exponentially. That is there exists $k \in \mathbb{R}^+$, $\nu \in (0,1)$:*

$$d(T^n\xi, T^n\eta) \le K\nu^n d(\xi,\eta).$$

Proof. Let $\Lambda_\delta = \{x \in X \mid x \text{ has atbale manifold of size } \delta\}$, by hypothesis

$$m(\cup_{k\in\mathbb{N}}\Lambda_{\frac{1}{k}}) = m(X).$$

Moreover, by definition, for each $\delta > 0$ and $x \in \Lambda_\delta$ there exists $n_\delta(x) \in \mathbb{N}$ such that

$$(T^n\xi,\, T^n\eta) < \frac{1}{2}d(\xi,\eta) \quad \forall \xi,\eta \in W^s_\delta(x),\ n > n_\delta(x).$$

Let us then define $\Lambda_{\delta,L} := \{x \in \Lambda_\delta \mid n_\delta(x) \le L\}$. Clearly $m(\cup_{L\in\mathbb{N}}\Lambda_{\delta,L}) = m(\Lambda_\delta)$ for each $\delta > 0$. By the above remarks we can choose δ, L such that $m(\Lambda_{\delta,L}) > 0$. We start by considering the reurn map $T_{\delta,L}$ to the set $\Lambda_{\delta,L}$. It is then natural to define the return times $\{n_j(x)\}$, $x \in \Lambda_{\delta,L}$, such that $T^{n_j(x)}x = T^j_{\delta,L}x$. Thus, for each $n \in \{n_j(x), \dots, n_{j+1}(x)-1\}$ holds

$$d(T^n\xi,\, T^n\eta) \le 2^{-j}d(\xi,\eta). \tag{3.4.12}$$

Lemma 3.4.7 *Given $U \subset X$ such that (3.4.12) holds, almost every $x \in U$ has a manifold $W_\delta(x)$ that contracts exponentially.*

Proof. Let $j_n(x) := \frac{1}{n} \sum_{i=1}^{n} \chi_U(T^i x)$, then (3.4.12) reads

$$d(T^n \xi, T^n \eta) \leq 2^{-n j_n(x)} d(\xi, \eta)$$

but Birkhoff Theorem implies that $j_n \to j^+$ almost everywhere and that $A_0 := \{x \in U \mid j^+(x) = 0\}$ has measure zero. Indeed, $A = \{x \in X \mid j^+(x) = 0\}$ is an invariant set and so we have

$$0 = m(\chi_A j_+) = m(\chi_{A \cap U}) = m(A_0).$$

Thus for almost all $x \in U$ there exists $\bar{n}$ such that $j_n(x) \geq \frac{1}{2} j^+(x)$ for all $n \geq \bar{n}$. From this it follows immediately that there exists $K > 0$:

$$d(T^n \xi, T^n \eta) \leq K 2^{-\frac{n}{2} j_n(x)} d(\xi, \eta)$$

□

□

The importance of the hyperbolic systems lies almost entirely in the following result.[7]

Theorem 3.4.8 (Pesin [58]) *Let (X, μ, T) be a Smooth Hyperbolic System, almost each point $x \in X$ has a local stable manifolds $W^s_\varepsilon(x)$, for some $\varepsilon > 0$, that is a manifold with the following properties:*

1. $x \in W^s_\varepsilon(x)$;
2. $\mathcal{T}_x W^s_\varepsilon(y) = E^s(y) \quad \forall y \in W^s(x)$;
3. *there exists $K > 0$, and $\nu \in (0, 1)$ such that, for each $y \in W^s_\varepsilon(x)$ $d(T^n x, T^n y) \leq K \nu^n d(x, y)$.*

and such manifolds are unique.

Note that, if T is invertible, the previous theorem applied to T^{-1} yield a local unstable manifold. Moreover, if $W^s(x)$ is a local stable manifold at x, then there must exist local stable manifolds at Tx as well (since $TW^s(x)$ is one of them). From now on we will assume that W^s are the local stable manifolds provided by Pesin Theorem, without entering in a real discussion concerning their size, which can be proven to enjoy some extremely mild form of uniformity.

We will not prove 3.4.8 in its full generality, but see 7 for a proof in a less general setting.

[7] To appreciate better the next theorem consider that the Oselec Theorem provides invariant distributions made of stable and unstable subspaces, then Pesin theorem essentially says that such distributions are integrable. In general, to integrate a distribution are necessary both regularities and geometrical properties (crf. Frobenius Theorem [11]) that, typically, are not verified (regularity) and very hard to verify (geometry) the present setting. Only the fact that the distribution is of a dynamical origin saves the day.

Definition 3.4.9 *At each point x for which the local stable manifold exists we can define the global stable manifold has*

$$\overline{W}^s(x) = \bigcup_{n=0}^{\infty} T^{-n} W^s(T^n x).$$

Remark 3.4.10 *One can then see that the general situation is not too dissimilar from the two dimensional automorphism of the torus, in particular, if the system is Anosov, then $\overline{W^s}$ turns out to be of infinite size. Yet, if the system is non-uniformly hyperbolic, then $\overline{W^s}$ can be arbitrarily short or it may be long in its own metric but wiggle so much as to be contained in a very small ball.*

The only other property that was really used in chapter 2 was the possibility to use the stable and unstable foliation as the basis for new coordinates. The relevant property, since we are dealing with measures, is that the change of coordinates does not substantially alter the measure: that is the new measure (obtained by the invariant one via the change of coordinates) is absolutely continuous with respect to the invariant one. If we want to extend the applicability of such an argument we need some analogous property. The right generalization turns out to be the so called *absolute continuity* . To describe it let us consider $x \in X$ with $W^s(x) \neq \emptyset$. Then, for a.e. such x, there exists a sufficiently small neighborhood $U(x)$ such that, for any two manifolds W_1, W_2 transversal to $W^s(x)$, the set $A = \{z \in W_1 \cap U(x) \mid W^s(z) \cap W_2 \neq \emptyset\}$ is not empty. In addition, if by μ_{W_i} we mean the measure induced on W_i by the Riemannian metric, $\mu_{W_1}(A) > 0$. Let us define $\phi : A \to W_2$ by $\phi(z) = W^s(z) \cap W_2$. We state the following in the simplest case.

Theorem 3.4.11 (Pesin [58]) *If the system (X, μ, T) is smooth, hyperbolic and μ is the Riemannian volume, then, for almost all $x \in X$ the measure $\phi_* \mu_{W_1}$ is absolutely continuous with respect to the measure μ_{W_2}.*

In our example we had the extreme case $\phi_* \mu_{W_1} = \mu_{W_2}$, yet it turns out that the above weaker property suffices to push the Hopf argument through, as we will see in chapters **??**, **??**. For the time being let us only quote the following general fact:

Theorem 3.4.12 ([58],[50]) *If the system (X, μ, T) is smooth, hyperbolic and μ is the Riemannian volume, then it has, at most, countably many ergodic components.*[8]

3.5 Comments on the non-smooth case

The results of the last section, although quite general, have a shortcoming: the smoothness requirements. As we will see in the following, systems that are quite natural both from the mathematical point of view and from the physical one are not smooth–typically they have discontinuities. In this section we will discuss a class of systems called *smooth systems with singularities*. Although the theory of such systems has been done in great generality, here we will give a restrictive definition, just sufficient for our later purposes. See the notes at the end of the chapter for information on more general settings.

[8] An ergodic component is a set which is the support of an invariant ergodic measure. The ergodic components of a measure are the measures associated to its ergodic decomposition. In other words μ can be written as the convex combination of at most countably many ergodic measures.

Definition 3.5.1 *By Smooth Dynamical System with singularities we mean a Dynamical Systems (X, T, μ), where*

- *X is the union of finitely many compact pieces X_i of $\mathbb{R}^n$, ∂X_i is the union of finitely many $n-1$ dimensional smooth manifolds.*
- *T is smooth outside a compact set $\mathcal{S}$. The singularity set $\mathcal{S}$ is the finite union of smooth $n-1$ dimensional manifolds with boundary $\mathcal{S}_i$, $\mathcal{S}_i \cap \mathcal{S}_j \not\subset \partial\mathcal{S}_i \cap \partial\mathcal{S}_j$ implies $i = j$. In addition, the boundary $\partial\mathcal{S}_i$ is the finite union of smooth $n-2$ dimensional manifolds.*
- *There exists $c_1, c_2 > 0$ such that*
$$\|D_x T\| + \|D_x^2 T\| \leq c_1 \mathit{dist}\,(x, \mathcal{S})^{-c_2}.$$
- *The measure μ is absolutely continuous with respect to Lebesgue.*

Remark 3.5.2 *Note that the fact that (X, T, μ) is a Smooth Dynamical System with singularities does not implies immediately that the same holds for (X, T^k, μ). The problem is that the map T can be very wild near the set $\mathcal{S}$, so it is not clear that the singularity set of T^k will satisfy our requirements. Nevertheless, in the examples we will consider, all the Dynamical System (X, T^k, μ) will always be Smooth Dynamical System with singularities.*

Remark 3.5.3 *We will call a smooth Dynamical System with singularities* invertible *if T^{-1} is densely defined and (X, T^{-1}, μ) is itself a smooth Dynamical System with singularities.*

Note that the above conditions imply the applicability of Oseledets Theorem.

3.5.1 Examples

Backer map

It is easy to check that the Backer map is a Smooth Dynamical System with singularities.

Discontinuous Arnold cat

If we consider $(\mathbb{R}^2, L, m)$ where

$$L \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & a \\ a & 1 + a^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{3.5.13}$$

with $a \notin \mathbb{Z}$, then it is not possible to project the system down to a torus preserving the continuity of the map. Yet, we can construct a discontinuous version of the Arnold cat.

Consider $\mathcal{M}_+ = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x + ay < 1;\ 0 \leq y < 1\}$ and $\mathcal{M}_- = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x < 1;\ 0 \leq ax - y < 1\}$. It is easy to see that, if Π is the projection from the universal cover $\mathbb{R}^2$ to the torus $\mathbb{T}^2$ ($\Pi\xi = \xi \mod 1$), then Π, restricted to $\mathcal{M}_\pm$, is one-one and onto. Moreover, $L\mathcal{M}_+ = \mathcal{M}_-$. This means that we can define $T : \mathbb{T}^2 \to \mathbb{T}^2$ by

$$T = \Pi L (\Pi|_{\mathcal{M}_+})^{-1}.$$

Of course T is discontinuous on $\mathcal{S}_+ := \partial\mathcal{M}_+$ and T^{-1} is discontinuous on $\mathcal{S}_- := \partial\mathcal{M}_-$. In addition, the Lebesgue measure is invariant and the map is hyperbolic since $DT = L$.

The question arises if there exists stable and unstable manifolds. A moment of thought shows that this is equivalent to the following question: there exist segments in the stable (unstable) direction such that their images in the future (past) never meet the discontinuity set $\mathcal{S}_+$ ($\mathcal{S}_-$)?

Let us analyze the unstable manifolds. Call $\mathcal{S}_\delta$ the δ neighborhood of $\mathcal{S}_+$. Consider a segment J centered at x, in the unstable direction, and suppose that $T^{-n}J \cap \mathcal{S}_+ \neq \emptyset$, then J cannot be the unstable manifold since its points do not have the same asymptotic trajectory in the past. Let $\lambda > 1$ the eigenvalue of L, then $T^{-n}J$ has total length $\lambda^{-n}|J|$, so the trajectory of x can be fairly close to $\mathcal{S}_+$ without having a problem. This discussion leads naturally to considering the set

$$G_\delta = \{x \in \mathbb{T}^2 \mid \text{dist } (T^{-n}x, \mathcal{S}_+) \geq \lambda^{-n}\delta\}.$$

On the one hand, it is clear that if $x \in G_\delta$, a segment in the unstable direction of size δ is indeed an unstable manifold. On the other hand, $m(G_\delta) \leq c\delta$. Thus almost all the points do have an unstable manifold of some positive size. This it is encouraging, yet it is clearly not sufficient to perform the Hopf argument. We will discuss this further in chapter **??**. For the time being it suffices to notice that what we have seen so far implies that the discontinuous Arnold cat has, at most, countably many ergodic components (see Problem counterg).

The interest of the Smooth Dynamical System with singularities is that most of the theory of the previous sections holds for this more general systems. Namely Theorems 3.4.8, 3.4.11, [45].

3.6 Flows

All what we have described so far has a rather straightforward generalization in the case of flows, yet some natural changes are called for.

To appreciate the problem let us consider a flow, on a compact Riemannian manifold, generated by a smooth non-zero vector field V. By definition $\frac{d}{dt}\phi^t|_{t=0} = V(x)$ and $d\phi^t V(x) = V(\phi^t x)$, thus $\lambda(x, V(x)) = 0$. This is a rather general fact: the Lyapunov exponents in the flow direction is zero. The only relevant exception is constituted by hyperbolic fixed points (think of the unstable equilibrium point of the pendulum) that, in the previous example, was ruled out by the assumption that the vector field be non zero. We will consider only such case.

Consequently a flow is hyperbolic if the tangent space is split in three transversal subspaces E^s, E^u, E^0, where E^0 is the flow direction and corresponds to a zero Lyapunov exponents.

Oseledec Theorem (Theorem 3.2.2) holds unchanged with (3.2.1) obviously replaced by

$$\int_X \| \log d\phi^t \| d\mu < \infty.$$

For a smooth flow coming from a non vanishing vector field Theorem 3.2.3 holds unchanged as well, the same for Theorems 3.4.8, 3.4.11 and 3.4.12.

The extension to non-smooth cases require a bit of care. we will do it only for the symplectic case (Hamiltonian flows with collisions), which is the focus of the following chapters.

3.6.1 Examples

Smooth flows with collisions

Let M be a smooth manifold with piecewise smooth boundary ∂M. We assume that the manifold M is equipped with a symplectic structure ω.[9] Given a smooth function H on M with *non vanishing* differential we obtain the non vanishing Hamiltonian vector field $F = \nabla_\omega H$ on M by $\omega(\nabla_\omega H, v) = dH(v)$. The vector field F is tangent to the level sets of the Hamiltonian $M^c = \{z \in M | H(z) = c\}$.

We distinguish in the boundary ∂M the regular part, ∂M_r, consisting of the points which do not belong to more than one smooth piece of the boundary and where the vector field F is transversal to the boundary. The regular part of the boundary is further split into "outgoing" part, ∂M_-, where the vector field F points outside the manifold M and the "incoming" part, ∂M_+, where the vector field is directed inside the manifold. Suppose that additionally we have a piecewise smooth mapping $\Gamma : \partial M_- \to \partial M_+$, called the collision map. We assume that the mapping Γ preserves the Hamiltonian, $H \circ \Gamma = H$, and so it can be restricted to each level set of the Hamiltonian.

We assume that all the integral curves of the vector field F that end (or begin) in the singular part of the boundary lie in a codimension 1 submanifold of M.

We can now define a flow $\Psi^t : M \to M$, called a flow with collisions, which is a concatenation of the continuous time dynamics Φ^t given by the vector field F, and the collision map Γ. More precisely a trajectory of the flow with collisions, $\Psi^t(x),\ x \in M$, coincides with the trajectory of the flow Φ^t until it gets to the boundary of M at time $t_c(x)$, the collision time. If the point on the boundary lies in the singular part then the flow is not defined for times $t > t_c(x)$ (the trajectory "dies" there). Otherwise the trajectory is continued at the point $\Gamma(\Psi^{t_c}x)$ until the next collision time, i.e., for $0 \le t \le t_c\left(\Gamma(\Psi^{t_c(x)}x)\right)$

$$\Psi^{t_c+t}x = \Phi^t \Gamma \Psi^{t_c} x.$$

We define a flow with collisions to be symplectic, if for the collision map Γ restricted to any level set M^c of the Hamiltonian we have

$$\Gamma^* \omega = \omega,$$

for some non vanishing function β defined on the boundary. More explicitly we assume that for every vectors ξ and η from the tangent space $T_z \partial M^c$ to the boundary of the level set M^c we have

$$\omega(D_z\Gamma\xi, D_z\Gamma\eta) = \omega(\xi, \eta).$$

We restrict the flow with collisions to one level set M^c of the Hamiltonian and we denote the resulting flow by Ψ_c^t. This flow is very likely to be badly discontinuous but we can expect that for a fixed time t the mapping Ψ_c^t is piecewise smooth, so that the derivative $D\Psi_c^t$ is well defined except for a finite union of codimension one submanifolds of M^c. We will consider only such cases.

The symplectic volume $\wedge^d \omega$ is clearly invariant for the flow, so will be the measure μ_c obtained by restricting the symplectic volume to the manifold M^c. Clearly for such an invariant measure all the trajectories that begin (or end) in the singular part of the boundary have measure zero. With respect to the measure μ_c the flow Ψ_c^t is a measurable flow in

[9] That is a non–degenerate closed antisymmetric two form.

the sense of Definition 1.1.3 and we obtain a measurable derivative cocycle $D\Psi_c^t : T_x M^c \to T_{\Psi_c^t x} M^c$. We can define Lyapunov exponents of the flow Ψ_c^t with respect to the measure μ_c, if we assume that

$$\int_{M^c} \log_+ \|D_x \Psi_c^t\| d\mu_c(x) < +\infty$$
$$\int_{\partial M_-^c} \log_+ \|D_y \Gamma\| d\mu_{cb}(y) < +\infty \tag{3.6.14}$$

(cf.[57]).

Problems

3.1 Prove that $\lambda(Tx, D_x T v) = \lambda(x, v)$.

3.2 Prove that $\lambda(x, v + w) \leq \max\{\lambda(x, v), \lambda(x, w)\}$ and $\lambda(x, \alpha v) = \lambda(x, v)$ for each $\alpha \in \mathbb{R}$, if they all exist. (Hint: Just apply the definition of LE and note that

$$\lambda(x, v + w) \leq \lim_{n\to\infty} \max\{\frac{1}{n} \log \|D_x T^n v\|, \frac{1}{n} \log \|D_x T^n w\|\}. \quad)$$

3.3 Assuming only that the LE are well defined a.e., prove that, if (X, T, μ) is ergodic, X is a d dimensional manifold and T a diffeomorphism, then there exists d numbers $\{\lambda_i\}$ such that the Lyapunov exponents $\lambda(x, v) \in \{\lambda_i\}$ a.e.. (Hint: For each $\alpha \in \mathbb{R}$ define $V_\alpha(x) := \{v \in \mathcal{T}_x X \mid \lambda(x, v) \leq \alpha\}$. By Problem 3.2 $V_\alpha(x)$ is a linear vector space and, by Problem 1 the distribution V_α is invariant. Then $d_\alpha(x) := \dim V_\alpha(x)$ is an invariant function, thus a.e. constant for each α. In addition, d_α is an increasing function of α and can assume only the values $\{0, \dots, d\}$. Thus there are at most $s \leq d$ $\{\alpha_j\}$ where d_α jumps. But this means that the LE are discrete. In fact, let $v \in V_\alpha(x) \backslash V_\beta(x)$, $\alpha > \beta$, then for each $w \in \text{span}\{v, V_\beta(x)\}$ it is easy to compute that $\lambda(x, w) = \lambda(x, v) > \beta$, which means: the LE is constant over $V_\alpha(x)$ apart for lower dimensional subspaces. In addition, we have a flag of subspaces $\{V_i\}_{i=0}^s$, $s \leq d$, such that $V_\alpha \in \{V_i\}_{i=0}^s$ for each $\alpha \in \mathbb{R}$. Hence, if $V_\alpha \supset V_i$ but $V_\alpha \not\supset V_{i+1}$ it must be $V_\alpha = V_i$, thus if $v \in V_\alpha$ but $v \notin V_{i-1}$ $\lambda(x, v) = \alpha_i$ where $\alpha_i = \inf\{\alpha \in \mathbb{R} \mid V_\alpha \supset V_i\}$.)

3.4 Show that, if T is invertible, $\{\lambda_i(x)\}$ is equal a.e. to $\{-\lambda_i^-(x)\}$ where $\{\lambda_i^-(x)\}$ are the LE of (X, T^{-1}, μ).

3.5 Show that

$$\lim_{n\to\infty} \frac{1}{n} \log |\det(D_x T^n)|$$

exists almost everywhere. (Hint: Apply BET.)

3.6 Let (X, T, μ) be a Dynamical Systems, X a compact Riemannian manifold and T a.e. differentiable. Suppose that there exists a one-dimensional distribution $E(x)$ such that $D_x T E(x) = E(Tx)$. Prove, without using Oseledets theorem, that for each $v \in E(x)$ the LE $\lambda(x, v)$ is well defined. (Hint: Let $v(x) \in E(x)$, $\|v(x)\| = 1$, then $D_x T v(x) = \alpha(x) v(Tx)$ and thus $D_x T^n v(x) = \prod_{i=1}^n \alpha(T^i x) v(T^i x)$. Then the result follows by the BET.)

3.7 Show that, if p is a fixed point, then $TW^s(p) \subset W^s(p)$ and $TW^u(p) \supset W^u(p)$.

3.8 Prove that the set $B^*_{\delta,c}$ in section 3.3 is closed with respect to the sup norm $\|u\|_\infty = \sup_{t\in[-\delta,\delta]} |u(t)|$.

3.9 Prove that the limit in (3.3.10) is well defined.

3.10 Prove that, in the setting of Theorem 3.3.2, the unstable manifold is unique. (Hint: This amounts to show that the set of points that are attracted to zero are exactly the manifolds constructed in Theorem 3.3.2. Use the local hyperbolicity to show that.)

3.11 Consider the Levowich map (3.3.11), show that, given the set of curves $\Gamma_{\delta,c} := \{\gamma : [-\delta,\, \delta] \to \mathbb{R}^2 \mid \gamma(t) = (t,\, u(t));\ \gamma(0) = 0;\ |u'(t)| \in [c^{-1}t,\, ct]\}$, it is possible to construct the map $\tilde{T} : \Gamma_{\delta,c} \to \Gamma_{\delta(1+c^{-1}\delta),\, c}$ in analogy with (3.3.4).

3.12 In the case of the previous problem show that for each $\gamma_i \in \Gamma_{\delta,c}$ holds $d(\tilde{T}\gamma_1,\, \tilde{T}\gamma_2) \le (1 - c\delta)d(\gamma_1,\, \gamma_2)$.

3.13 Show that for the Levowich map zero has a unique unstable manifold. (Hint: grow the manifolds, that is, for each $n > 1$ define $\delta_n := \frac{\rho}{n}$. Show that one can choose ρ such that $\delta_{n-1} \ge \delta_n(1 + c^{-1}\delta_n)$. according to Problem 11 it follows that $\tilde{T} : \Gamma_{\delta_n,c} \to \Gamma_{\delta_{n-1},c}$. Moreover,

$$d(\tilde{T}^{n-1}\gamma_1,\, \tilde{T}^{n-1}\gamma_2) \le \prod_{i=1} n(1 + c^{-1}\delta_i)d(\gamma_1,\, \gamma_2).$$

Finally, show that, setting $\gamma_n(t) = (0, t) \in \Gamma_{\delta_n,c}$, the sequence $\tilde{T}^{n-1}\gamma_n$ is a Cauchy sequence that converges to a curve in $\Gamma_{1,c}$ invariant under $\tilde{T}$.)

3.14 Prove that Oselec Theorem can be applied to smooth Dynamical Systems with singularities. (Hint: Just check the (3.2.1) is satisfied.)

3.15 Consider the discontinuous Arnold cat. That is, the map $T : [0, 1]^2 \to [0, 1]^2$ defined by $T(x, y) = \begin{pmatrix} 1 + a^2 & a \\ a & 1 \end{pmatrix}$ mod 1 with $a \notin \mathbb{Q}$. Prove that $(T, [0, 1]^1, \text{Leb})$ has, at most, countably many ergodic components. (Hint: Imitate the Hopf argument perform in chapter 2. First of all notice that, by Fubini theorem, almost all segment has a full measure of point on which f^+ and f^- are well defined and equal. Thus almost all points have a stable manifold of some length with almost all point on it with unstable manifold of some length. This allows immediately to construct that almost all points belong to an invariant set of positive measure on which the map is ergodic. Clearly there can be at most countably many such sets.)

3.16 define cocycles (Hint: The derivative of the flow with collisions can be also naturally factored onto the quotient of the tangent bundle TM^c of M^c by the vector field F, which we denote by $\widehat{T}M^c$. Note that for a point $z \in \partial M^c$ the tangent to the boundary at z can be naturally identified with the quotient space. We will again denote the factor of the derivative cocycle by

$$A^t(x) : \widehat{T}_x M^c \to \widehat{T}_{\Psi^t_c x} M^c.$$

We will call it the transversal derivative cocycle. If the derivative cocycle has well defined Lyapunov exponents then the transversal derivative cocycle has also well defined

Lyapunov exponents which coincide with the former ones except that one zero Lyapunov exponent is skipped.)

Notes

The point of view of obtaining invariant manifold by fixed points arguments is brought to in extreme consequences in the [39], where a rather general theory of invariant foliations is presented.

It is interested that the dynamics near an hyperbolic fixed point can be conjugated (continuously) to the linear part, this is called the Hartman-Grobman theorem, see [43].

The theory of foliations for piecewise continue maps is developed in great generality in [45].

CHAPTER 4

Hyperbolicity and the magic of cones

In chapter 2 we have taken advantage of the hyperbolic structure of the map in a very explicit manner, yet non linear maps may be hyperbolic but this cannot be seen by naively analyzing their derivative. Hence the necessity to have a tool to establish the hyperbolicity of a given system.

Our next task is to understand a general approach which allows to establish when the Lyapunov exponents are different from zero almost everywhere.

4.1 The two dimensional case

We start by dealing with the area preserving two dimensional case in order to explain the basic idea.

Theorem 4.1.1 (Wojtkowski [74]) *Let (X, μ, T) be a dynamical system where X is a compact two dimensional Riemannian manifold, μ is the Riemannian volume,*[1] *T a diffeomorphism of X and*

$$\int_X \log \|DT\| d\mu < \infty.$$

If there exists a measurable family of convex two sided cones $\mathcal{C}(x) \subset \mathcal{T}_x X$ such that, for almost all $x \in X$, there exists $n \in \mathbb{N}$ with the property[2]

$$D_x T^n \mathcal{C}(x) \subset int(\mathcal{C}(T^n x)) \cup \{0\},$$

then the Lyapunov exponents are different from zero almost surely.

A system with a family of cones satisfying the hypotheses of the Theorem 4.1.1 is called *eventually strictly monotone.*[3]

[1] In fact, it is the symplectic not the Riemannian structure that matters; yet, in two dimension, there is not much difference. On the contrary, in section 4.2 this difference will be made explicit.

[2] A "(two sided) cone," in a linear space, is a set such that if ξ belongs to the set then $\lambda\xi$ belongs to the set for each $\lambda \in \mathbb{R}$. In addition, we require that the set is closed, has open interior and its complement is non void. Actually, this last conditions could be relaxed for x in a zero measure set without changing the following proof (see footnote 6 at page 84).By "two sided convex cone" we mean that each half cone is a convex set. Notice that in $\mathbb{R}^2$ a such a cone is defined uniquely by the two edges. By measurable we mean that the functions from X to the unit vectors, in the direction of the edges, are measurable.

[3] In fact, in the case of billiards a similar situation is called *sufficiency* but I find the above terminology more appropriate.

PROOF. Let $x \in X$ and $n \in \mathbb{N}$ such that

$$D_x T^n \mathcal{C}(x) \subset \operatorname{int}(\mathcal{C}(T^n x)) \cup \{0\}.$$

The first thing to notice is that it is possible to make an orientation preserving change of coordinates (i.e., a change of coordinates via a matrix with positive determinant) both in $\mathcal{T}_x X$ and in $\mathcal{T}_{T^n x} X$ such that, in the new coordinates, $\mathcal{C}(x)$ and $\mathcal{C}(T^n x)$ become the standard cone $\mathcal{C}_+ = \{v \in \mathbb{R}^2 \mid v_1 v_2 \geq 0\}$ and the Riemannian structures–the scalar product and the volume–are the standard ones (see Problem 4.5, Problem 4.6). Viewed in this coordinates $D_x T^n$ becomes a two by two matrix with determinant equal to one, that maps $\mathcal{C}_+$ strictly into itself. Note that, since the cone family is measurable, the change of coordinates depends measurably by x.

To continue it is necessary to study a bit the general properties of the matrices enjoying the above mentioned properties. Notice that if we define a quadratic form $Q : \mathbb{R}^2 \to \mathbb{R}$ by $Q(v) = v_1 v_2$, then $\mathcal{C}_+ = \{v \in \mathbb{R}^2 \mid Q(v) \geq 0\}$, so our task is to study the two by two matrices L with $\det(L) = 1$ and such that $Q(v) \geq 0$, $v \neq 0$, implies $Q(Lv) > 0$.[4]

Algebraic considerations

Let $v = (1, u)$ with $u \in \mathbb{R}^+$, which implies $v \in \mathcal{C}_+$, and

$$L = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $\det(L) = 1$ (note that this is equivalent to L symplectic, see section 4.2). Then, for each $u \geq 0$, we must have

$$0 < Q(Lv) = ac + (ad + bc)u + bdu^2, \tag{4.1.1}$$

Setting $u = 0$, it must be $ac > 0$. On the other hand, since $(0, 1) \in \mathcal{C}_+$ and $L(0, 1) = (b, d)$, it must be $bd > 0$. Finally, if we compute the quadratic polynomial in its minimum $u_0 = -\frac{ad+bc}{2bd}$ we get, calling $v_0 = (1, u_0)$,

$$Q(Lv_0) = -\frac{1}{4bd} < 0.$$

The above relation is possible only if $v_0 \notin \mathcal{C}_+$, which implies $u_0 < 0$ or $ad + bc > 0$, that is $ad > \frac{1}{2}$. Collecting the above results it follows that all the elements of the matrix L must be different from zero and, in addition, they must have the same sign. Since $Q(v) = Q(-v)$, without loss of generality we can assume them to be all positive.

The next step is to define some measure of expansion for a strictly monotone matrix. A natural quantity to consider is:

$$\sigma(L) = \inf_{v \in \operatorname{int}(\mathcal{C}_+)} \sqrt{\frac{Q(Lv)}{Q(v)}}.$$

Choosing again $v = (1, u)$, it follows

$$\frac{(a + bu)(c + du)}{u} \geq ad + bc = 1 + 2bc > 1, \tag{4.1.2}$$

[4] We will call such matrices *Strictly monotone*.

thus $\sigma(L) \geq 1$.

Moreover, given two monotone matrices L_1, L_2, we have

$$\sigma(L_1 L_2) = \inf_{v \in \mathrm{int}(\mathcal{C}_+)} \sqrt{\frac{Q(L_1 L_2 v)}{Q(L_2 v)}} \sqrt{\frac{Q(L_2 v)}{Q(v)}} \geq \sigma(L_1)\sigma(L_2). \tag{4.1.3}$$

An interesting fact, that follows immediately from (4.1.2), is that $\sigma(L) > 1$ if and only if L is strictly monotone.

Measure theoretical considerations

The point of measuring the expansion via the Q-form is due to the following Lemma.[5]

Lemma 4.1.2 *If the Dynamical System (X, T, μ), X a two dimensional Riemannian manifold, T differentiable a.e. and μ the Riemannian volume, is eventually strictly monotone, then*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \sigma(D_x T^n) > 0 \quad \mu\text{-a.e.}.$$

Proof. Let $\nu : X \to \mathbb{R}^+ \cap \{\infty\}$ be defined by

$$\nu(x) := \liminf_{n\to\infty} \frac{1}{n} \ln \sigma(D_x T^n).$$

Then

$$\begin{aligned}
\nu(T^{-1}x) &= \liminf_{n\to\infty} \frac{1}{n} \ln \sigma(D_{T^{-1}x} T^n) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \{\ln \sigma(D_x T^{n-1}) + \ln \sigma(D_{T^{-1}x} T)\} \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \ln \sigma(D_x T^n) = \nu(x),
\end{aligned} \tag{4.1.4}$$

where we have used (4.1.3) and the fact that $\sigma(D_\xi T) \geq 1$ by monotonicity.

Let $A_0 = \{x \in X \mid \nu(x) = 0\}$, to prove the Lemma it suffices to show that $\mu(A_0) = 0$. To this end note that $TA_0 \subset A_0$, since (4.1.4) implies $\nu(Tx) \leq \nu(x)$. Then, consider $\Lambda = \cup_{n\in\mathbb{N}} T^{-n} A_0$, clearly the $\Lambda \supset A_0$ is an invariant set and

$$\mu(\Lambda \backslash A_0) = \mu(\cup_{n\in\mathbb{N}} T^{-1} A_0 \backslash A_0) \leq \sum_{n=0}^{\infty} [\mu(T^{-n} A_0) - \mu(A_0)] = 0. \tag{4.1.5}$$

Consequently, if we suppose that $\mu(A_0) > 0$, then $\mu(\Lambda) > 0$. Therefore, for each $m \in \mathbb{N}$,

$$0 = \int_{A_0} \nu(x)\mu(dx) \geq \int_{A_0} \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{[\frac{n}{m}]-1} \ln \sigma(D_{T^{im}x} T^m),$$

[5] It is interesting to remark that the next Lemma, together with the results of section 4.3, imply the existence of the stable and unstable distribution (see Examples 4.3.1-Cones and Q-forms). Thus, since X is two dimensional, the existence a.e. of the L.E. follows as in Problem 3.6 without invoking Oseledets Theorem. The use of Oseledets Theorem is instead necessary in higher dimensions.

where we have used (4.1.3) again. At this point we would like to use BET, yet we face a technical problem: we do not know if $\ln \sigma(D_x T^m)$ is integrable. Nevertheless, we are not interested in large values of $\ln \sigma(D_x T^m)$. It is then natural to define

$$\varphi_m(x) = \min\{\ln \sigma(D_x T^m),\, 1\}.$$

Now $\varphi_m \in L^\infty(X, \mu)$, thus the ergodic average $\varphi_m^+ \in L^\infty(X, \mu)$; hence, remembering (4.1.5),

$$\begin{aligned} 0 \geq \int_{A_0} \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{[\frac{n}{m}]-1} \varphi_m(T^{im}x) &= \frac{1}{m} \int_{A_0} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi_m(T^{im}x) \\ = \frac{1}{m} \int_{A_0} \varphi_m^+(x) &= \frac{1}{m} \int_\Lambda \varphi_m^+ = \frac{1}{m} \int_\Lambda \varphi_m. \end{aligned} \tag{4.1.6}$$

That is $\varphi_m = 0$ a.e. in Λ. But this is a contradiction since, calling $B_m = \{x \in X \mid \sigma(D_x T^m) > 1\}$, the definition of eventually strictly monotone is equivalent to $\mu(\cup_{m\in\mathbb{N}} B_m) = \mu(X)$. Therefore there must exists an $m \in \mathbb{N}$ such that $\mu(\Lambda \cap B_m) > 0$, which implies

$$\int_\Lambda \varphi_m \geq \int_{\Lambda\cap B_m} \varphi_m > 0,$$

whereby contradicting (7.1.1). □

The relevance of what we have seen up to now for the estimation of the Lyapunov exponents depends on the trivial inequality

$$\|v\|^2 \geq 2Q(v). \tag{4.1.7}$$

The only real problem left is that, due to our change of variable to put the cones into their standard form, the euclidean norm $\|\cdot\|$ in the new variables no longer correspond to the original norm in X (let us call such original norm, at the point x, $\|\cdot\|_x$). Nevertheless, the two norms must be equivalent by construction,[6] hence there must exists an everywhere strictly positive measurable function $a(x) \leq 1$ such that, for each $v \in \mathcal{T}_x X$

$$a(x)^{-1}\|v\| \geq \|v\|_x \geq a(x)\|v\|.$$

Let us introduce the set $A(\varepsilon) = \{x \in X \mid a(x) > \varepsilon\}$, clearly $\cup_{\varepsilon>0} A(\varepsilon)$ has full measure. Now Poincaré theorem implies, if $\mu(A(\varepsilon)) \neq 0$, that almost all points in $A(\varepsilon)$ return to $A(\varepsilon)$ infinitely often. Let $x \in A(\varepsilon)$ be one of such points, then there exists a sequence n_k such that $T^{n_k}x \in A(\varepsilon)$ for each $k \in \mathbb{N}$.

Accordingly, for each $v \in \operatorname{int}(\mathcal{C}(x))$, $\|v\|_x = 1$, and $m \in \mathbb{N}$ holds

$$\begin{aligned} \lambda(x,\, v) = \lim_{n\to\infty} \frac{1}{n} \log \|D_x T^n v\|_{T^n x} &\geq \liminf_{k\to\infty} \frac{1}{n_k} \log \|D_x T^{n_k} v\|_{T^{n_k} x} \\ &\geq \liminf_{k\to\infty} \frac{1}{n_k} \log \|D_x T^{n_k} v\| + \liminf_{k\to\infty} \frac{1}{n_k} \log a(T^{n_k}x) \\ &\geq \liminf_{n\to\infty} \frac{1}{n} \log \|D_x T^n v\| - \lim_{k\to\infty} \frac{1}{n_k} \log \varepsilon^{-1} \end{aligned}$$

[6] If we admit that $\mathcal{C}(x)$ can have an empty interior on a set of zero measure in X–see footnote 2 at page 81–then the two norms would be equivalent only almost everywhere. Nevertheless, this does not change the proof: call X_1 the incriminated set, then $X_2 := \cup_{n\in\mathbb{Z}} T^n X_1$ has also zero measure and it is an invariant set. We can then discard such a set and work on its complement without any other change in the following.

and, since $Q(v) \neq 0$ and by (4.1.7),

$$\begin{aligned}\lambda(x,\, v) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \sqrt{\frac{Q(D_x T^n v)}{Q(v)}} \\ &\geq \liminf_{n\to\infty} \frac{1}{n} \log \sigma\left(D_x T^n\right) > 0,\end{aligned}$$

due to Lemma 4.1.2.

We have then seen that, for almost every $x \in X$ and $v \in Int\,\mathcal{C}(x)$, $\lambda(x,\, v) \neq 0$. Noticing that the Lyapunov exponents of T are given by minus the Lyapunov exponents of T^{-1} (seeProblem 3.4 and vicinity). Thus, by Oseledets Theorem [57] (or see section 4.4), almost every point must have a vector $v_-(x)$ such that $\lambda(x, v_-(x)) < 0$. Obviously, given any vector $v \in Int\,\mathcal{C}(x)$ it follows $\lambda(x, \alpha v + \beta v_-(x)) = \lambda(x, v)$, provided $\alpha \neq 0$, and this concludes the story. □

Remark 4.1.3 *The measurability assumption is a very weak hypothesis but cannot be eliminated. Indeed, if one constructs a cone family along the trajectories it can easily be made strictly monotone. Hence, if the system has zero Lyapunov exponents such a cone family cannot be measurable (see Problem 4.2).*

The above theorem provides us with a very powerful instrument to establish hyperbolicity for a given dynamical system.

To see how it works let us consider some simple examples.

4.1.1 Examples

Linear automorphysms of the Torus

Consider the matrix

$$L = \begin{pmatrix} 1 & a \\ a & 1+a^2 \end{pmatrix}$$

with $a \in \mathbb{N}$ and the standard cone $\mathcal{C}_+ = \{(v, v) \in \mathbb{R}^2 \mid uv \geq 0\}$. Then

$$L\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u + av \\ au + (1+a^2)v \end{pmatrix}$$

shows that $\mathcal{C}_+$ is strictly monotone for L. Of course, this is a rather silly example since it is completely obvious that the map is hyperbolic, the next example is a little less trivial.

Perturbations of linear automorphysms of the Torus

Consider a diffeomorphism $\Phi : \mathbb{T}^2 \to \mathbb{T}^2$ such that

$$\|(D\Phi - \mathbb{1})L\| < 1; \tag{4.1.8}$$

where L is defined as in the previous example, then the map T defined by $Tx := \Phi(Lx)$ is hyperbolic. To see this write

$$DT\begin{pmatrix} u \\ v \end{pmatrix} = L\begin{pmatrix} u \\ v \end{pmatrix} + (D\Phi - \mathbb{1})L\begin{pmatrix} u \\ v \end{pmatrix} := \begin{pmatrix} u + av \\ au + (1+a^2)v \end{pmatrix} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

but $\max\{|\alpha|, |\beta|\} \leq \|(D\phi - \mathbb{1})L\|\|(u,v)\| \leq u + v$. Thus $\mathcal{C}_+$ is strictly monotone for DT.

It is interesting to notice that, already for this simple example, it would be not immediately clear how to establish hyperbolicity without using a cone language. In addition, remark that the full strength of Wojtkowski theorem it is not used here–since the cone family is strictly monotone.

Levowich map

Let us consider the map $T : \mathbb{T}^2 \to \mathbb{T}^2$ defined by[7]

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - \sin x + y \\ x - \sin x + y \end{pmatrix},$$

It is immediate to verify that

$$D_{(x,y)}T = \begin{pmatrix} 2 - \cos x & 1 \\ 1 - \cos x & 1 \end{pmatrix}$$

Since $\det(DT) = 1$, it follows that $(\mathbb{T}^2, T, m)$ is a Dynamical Systems. In addition, $D_x T\mathcal{C}_+ \subset \mathcal{C}_+$ strictly, apart from the zero measure set $\{x = 0\}$, so Theorem 4.1.1 applies.

4.2 Higher dimension–an overview

The difficulties in extending to higher dimensions the previous results stem mainly from the large variety of possible cone shapes in higher dimension. It is far from obvious how to relate monotone properties of a given cone to the behavior of the Lyapunov exponents. One possible way is to generalize the approach based on quadratic forms. We will comment on this possibility in section **??**. Yet, in the special case of *Symplectic Systems*, it is possible to develop a very rich theory which is astonishingly similar to the two dimensional one. Here we give a brief insight into this theory, but see [53], [54] and [74] for a much more detailed account.

Definition 4.2.1 *By symplectic Systems we mean a Dynamical System (X, T, μ) where X is a symplectic manifold,*[8] *μ is the symplectic volume*[9] *and T is a symplectic map.*[10]

Clearly one can also define a Symplectic Systems in continuous time, we have already seem the typical example: Hamiltonian systems (see Examples 1.1.1).

For the convenience of the reader we will present here some of the material from [76] and [53].

Let $\mathcal{W}$ be a linear symplectic space of dimension $2d$ with the symplectic form ω. For instance we call $\mathcal{W} = \mathbb{R}^d \times \mathbb{R}^d$ the standard linear symplectic space if

$$\omega(w_1, w_2) = \langle \xi^1, \eta^2 \rangle - \langle \xi^2, \eta^1 \rangle, \tag{4.2.9}$$

[7]Note that here our torus has the periodicity of 2π instead than one as in the previous examples, this is just to have simpler formulae; the reader can easily reformulate the problem on the torus $\mathbb{R}^2 \mod 1$.

[8]A symplectic manifold is a smooth manifold of even dimensions together with a symplectic form. By *symplectic form* we mean an antisymmetric differential two form ω which is close, see [3] for more details.

[9]Given a symplectic form ω on a manifold of dimension $2d$ the $2d$ form $\wedge^d \omega$ is a volume form: the symplectic volume.

[10]A map is symplectic if is conserves the symplectic two form ω, that is, for each $x \in X$ and vectors $v, w \in \mathcal{T}_x X$, holds $\omega(D_x Tv, D_x Tw) = \omega(v, w)$.

where $w_i = (\xi^i, \eta^i)$, $i = 1, 2$, and $\langle \xi, \eta \rangle = \xi_1\eta_1 + \cdots + \xi_d\eta_d$.

The symplectic group $Sp\,(d, \mathbb{R})$ is the group of linear maps of $\mathcal{W}$ ($2d \times 2d$ matrices if $\mathcal{W} = \mathbb{R}^d \times \mathbb{R}^d$) preserving the symplectic form i.e., $L \in Sp\,(d, \mathbb{R})$ if

$$\omega(Lw_1, Lw_2) = \omega(w_1, w_2) \tag{4.2.10}$$

for every $w_1, w_2 \in \mathcal{W}$.

By definition a Lagrangian subspace of a linear symplectic space $\mathcal{W}$ is a d-dimensional subspace on which the restriction of ω is zero (equivalently it is a maximal subspace on which ω vanishes).

Definition 4.2.2 *Given two transversal Lagrangian subspaces V_1 and V_2 we define the sector between V_1 and V_2 by*

$$\mathcal{C} = \mathcal{C}\,(V_1, V_2) = \{w \in \mathcal{W} \mid \omega(v_1, v_2) \geq 0 \textit{ for } w = v_1 + v_2, v_i \in V_i, i = 1,\, 2\}$$

Equivalently, if we define the quadratic form associated with an ordered pair of transversal Lagrangian subspaces,

$$\mathcal{Q}(w) = \omega(v_1, v_2)$$

where $w = v_1 + v_2$, is the unique decomposition of w with the property $v_i \in V_i, i = 1, 2$, then we have

$$\mathcal{C} = \{w \in \mathcal{W} \mid \mathcal{Q}(w) \geq 0\}.$$

In the case of the standard symplectic space, $V_1 = \mathbb{R}^d \times \{0\}$ and $V_2 = \{0\} \times \mathbb{R}^d$ we get

$$\mathcal{Q}\left((\xi, \eta)\right) = \langle \xi, \eta \rangle$$

and

$$\mathcal{C}_+ = \{(\xi, \eta) \in \mathbb{R}^d \times \mathbb{R}^d \mid \langle \xi, \eta \rangle \geq 0\}.$$

We will refer to this $\mathcal{C}_+$ as the standard sector. Since any two pairs of transversal Lagrangian subspaces are symplectically equivalent (see Problem 4.18) we may consider only this case without any loss of generality.

It is natural to ask if a sector determines uniquely its sides. It is not a vacuous question since, for $d > 1$, there are many Lagrangian subspaces in the boundary of a sector. The answer is positive.

Proposition 4.2.3 *For two pairs of transversal Lagrangian subspaces V_1, V_2 and V_1', V_2' if*

$$\mathcal{C}\,(V_1, V_2) = \mathcal{C}\left(V_1', V_2'\right)$$

then

$$V_1 = V_1' \quad \textit{and} \quad V_2 = V_2'.$$

Moreover V_1 and V_2 are the only isolated Lagrangian subspaces contained in the boundary of the sector $\mathcal{C}\,(V_1, V_2)$.

Based on the notion of the sector between two transversal Lagrangian subspaces (or the quadratic form $\mathcal{Q}$) we define two monotonicity properties of a linear symplectic map. By $int\,\mathcal{C}$ we denote the interior of the sector, i.e.,

$$int\,\mathcal{C} = \{w \in \mathcal{W} | \mathcal{Q}(w) > 0\}.$$

Definition 4.2.4 *Given the sector $\mathcal{C}$ between two transversal Lagrangian subspaces we call a linear symplectic map L monotone if*

$$L\mathcal{C} \subset \mathcal{C}$$

and strictly monotone if

$$L\mathcal{C} \subset int\,\mathcal{C} \cup \{0\}.$$

A very useful characterization of monotonicity is given in the following

Proposition 4.2.5 *L is (strictly) monotone if and only if $\mathcal{Q}(Lw) \geq \mathcal{Q}(w)$ for every $w \in \mathcal{W}$ ($\mathcal{Q}(Lw) > \mathcal{Q}(w)$ for every $w \in \mathcal{W}$, $w \neq 0$). In particular, $Q(Lw) = Q(w)$, that is, L is a Q-isometry iff $L\mathcal{C} = \mathcal{C}$.*

The fact that monotonicity implies the increase of the quadratic form defining the cone is a manifestation of a very special geometric structure of a sector and does not hold for cones defined by general quadratic forms.

Proposition 4.2.6 *A monotone map L is strictly monotone if and only if*

$$LV_i \subset int\,\mathcal{C} \cup \{0\},\ i = 1, 2.$$

For the proofs of the above facts see [76] and [54].

Remark 4.2.7 *Proposition 4.2.3 and 4.2.6 are trivial in the two dimensional case. As already notice, proposition 4.2.5 follows, in the two dimensional case, by 4.1.2.*

The relevance of the above discussion is the possibility to extend Theorem 4.1.1 to the present setting.

Theorem 4.2.8 (Wojtkowski [74]) *Let (X, μ, T) be a dynamical system where X is the finite union of Symplectic Manifolds, μ is the symplectic volume, T an invertible almost everywhere differentiable symplectic map of X and*

$$\int_X \log \|DT\| d\mu < \infty.$$

If there exists a measurable, a.e. non degenerate, eventually strictly invariant family of sectors then the Lyapunov exponents are different from zero almost surely.

PROOF. The proof follows the one of Theorem 4.1.1 where the algebraic considerations are replaced by Propositions 4.2.3, 4.2.5, 4.2.6, while the measure theoretical part is exactly the same. □

4.2.1 Examples

Linear symplectic maps

We will consider the following generalization of the Arnold cat. Let us consider the Dynamical Systems $(\mathbb{T}^{2d}, T, m)$, where m is the Lebesgue measure and $Tx = Lx \mod 1$, with the following matrix L

$$L = \begin{pmatrix} \mathbb{1} & \mathbb{1} \\ M & \mathbb{1} + M \end{pmatrix}$$

where $M > 0$ and $M_{ij} \in \mathbb{Z}$ (see Problem 4.8 for a more concrete examples and Problem 4.15 to realize how general the example is). Then the system is symplectic and strictly monotone with respect to the standard sector, thus this Dynamical Systems is hyperbolic.

4.3 Metrics and cones

This section is devoted to a little digression on semi–metrics that can be associated to one sided cones convex. This is a vast field, here we will consider only few basic facts.

There is a very geometric approach to this: consider the projectivization of the cone (that is the set of equivalence classes with respect to the equivalence relation $\sim$ defined by $v \sim w$ iff there exists $\lambda \in \mathbb{R}^+$ such that $v = \lambda w$) whereby obtaining a convex set in the projective space and then use the associated projective metric [25]. We will use a more direct, yet equivalent, approach (see Problem 4.21, Problem 4.22 and Problem 4.23 for further informations on the above point of view and the connection with the following). Hopefully, the reader will excuse the setting which is a bit abstract in order to be applicable to some unexpected situations.

We start by illustrating some results in lattice theory originally due to Garrett Birkhoff. For more details see [10], and [56] for a recent overview of the field. Consider a topological vector space $\mathbb{V}$ with a partial ordering "$\preceq$," that is a vector lattice.[11] We require the partial order to be *continuous*, i.e. given $\{v_n\} \in \mathbb{V}$ $\lim_{n\to\infty} v_n = v$, if $v_n \succeq w$ for each n, then $v \succeq w$. We call such vector lattices *integrally closed.*[12]

We define the closed convex cone[13] $\mathcal{C} = \{v \in \mathbb{V} \mid v \neq 0,\, v \succeq 0\}$ (hereafter, the term "closed cone" $\mathcal{C}$ will mean that $\mathcal{C} \cup \{0\}$ is closed). Conversely, given a closed convex cone $\mathcal{C} \subset \mathbb{V}$, enjoying the property $\mathcal{C} \cap -\mathcal{C} = \emptyset$, we can define an order relation by (see Problem 4.20)

$$v \preceq w \iff w - v \in \mathcal{C} \cup \{0\}.$$

Henceforth, each time that we specify a convex cone we will assume the corresponding order relation and vice versa. The reader must therefore be advised that "$\preceq$" will mean different things in different contexts.

It is then possible to define a projective metric Θ (Hilbert metric),[14] in $\mathcal{C}$, by the construction:

$$\begin{aligned}
\alpha(v,\, w) &= \sup\{\lambda \in \mathbb{R}^+ \mid \lambda v \preceq w\}\\
\beta(v,\, w) &= \inf\{\mu \in \mathbb{R}^+ \mid w \preceq \mu v\}\\
\Theta(v,\, w) &= \log\left[\frac{\beta(v,\, w)}{\alpha(v,\, w)}\right]
\end{aligned} \tag{4.3.11}$$

where we take $\alpha = 0$ and $\beta = \infty$ if the corresponding sets are empty.

[11] We are assuming the partial order to be well behaved with respect to the algebraic structure: for each $v, w \in \mathbb{V}$ $v \succeq w \iff v - w \succeq 0$; for each $v \in \mathbb{V}$, $\lambda \in \mathbb{R}^+\backslash\{0\}$ $v \succeq 0 \Longrightarrow \lambda f \succeq 0$; for each $v \in \mathbb{V}$ $v \succeq 0$ and $v \preceq 0$ imply $v = 0$ (antisymmetry of the order relation).

[12] To be precise, in the literature "integrally closed" is used in a weaker sense. First, $\mathbb{V}$ does not need a topology. Second, it suffices that for $\{\alpha_n\} \in \mathbb{R}$, $\alpha_n \to \alpha$; $v, w \in \mathbb{V}$, if $\alpha_n v \succeq w$, then $\alpha v \succeq w$. Here we will ignore these and other subtleties: our task is limited to a brief account of the results relevant to the present context.

[13] Attention!!: here, by "cone," we mean any set such that, if v belongs to the set, then λv belongs to it as well, for each $\lambda > 0$. The reason for this change in the definition of cone is that two sided cones, viewed as sets, are never convex, while convexity plays a central rôle in the following. As we will see in Examples 4.3.1–Cones and Q-forms this change in definition does not limit the applicability of the present theory to the cones introduced in the previous section.

[14] In fact, we define a semi–metric, since $v \sim w \Rightarrow \Theta(v,\, w) = 0$. The metric that we describe corresponds to the conventional Hilbert metric on $\widetilde{\mathcal{C}}$, the quotient of $\mathcal{C}$ with respect to the relation "$\sim$".

Lemma 4.3.1 *The function Θ is a semi-metric in $\mathcal{C}$.*

PROOF. Clearly $\Theta(v,w) =$ implies $v = \lambda w$ for some $\lambda \in \mathbb{R}^+$, also $\Theta(v,w) = \Theta(w,v)$ and the tringle inequlity can be easily checked. □

The importance of the previous constructions is due, in our context, to the following theorem.

Theorem 4.3.2 *Let $\mathbb{V}_1$, and $\mathbb{V}_2$ be two integrally closed vector lattices; $L : \mathbb{V}_1 \to \mathbb{V}_2$ a linear map such that $L(\mathcal{C}_1) \subset \mathcal{C}_2$, for two closed convex cones $\mathcal{C}_1 \subset \mathbb{V}_1$ and $\mathcal{C}_2 \subset \mathbb{V}_2$ with $\mathcal{C}_i \cap -\mathcal{C}_i = \emptyset$. Let Θ_i be the Hilbert metric corresponding to the cone $\mathcal{C}_i$. Setting $\Delta = \sup_{v,\, w \in L(\mathcal{C}_1)} \Theta_2(v,\, w)$ we have*

$$\Theta_2(Lv,\, Lw) \leq \tanh\left(\frac{\Delta}{4}\right) \Theta_1(v,\, w) \qquad \forall v,\, w \in \mathcal{C}_1$$

($\tanh(\infty) \equiv 1$).

PROOF. Let $v,\, w \in \mathcal{C}_1$. On the one hand if $\alpha := \alpha(v,w) = 0$ or $\beta := \beta(v,w) = \infty$, then the inequality is obviously satisfied. On the other hand, if $\alpha \neq 0$ and $\beta \neq \infty$, then

$$\Theta_1(v,\, w) = \ln \frac{\beta}{\alpha}$$

where $\alpha v \preceq w$ and $\beta v \succeq w$, since $\mathbb{V}_1$ is integrally closed. Notice that $\alpha \geq 0$, and $\beta \geq 0$ since $v \succeq 0$, $w \succeq 0$. If $\Delta = \infty$, then the result follows from $\alpha Lv \preceq Lw$ and $\beta Lv \succeq Lw$. If $\Delta < \infty$, then, by hypothesis,

$$\Theta_2\left(L(w - \alpha v),\, L(\beta v - w)\right) \leq \Delta$$

which means that there exist $\lambda,\, \mu \geq 0$ such that

$$\begin{aligned} \lambda L(w - \alpha v) &\preceq L(\beta v - w) \\ \mu L(w - \alpha v) &\succeq L(\beta v - w) \end{aligned}$$

with $\ln \frac{\mu}{\lambda} \leq \Delta$. The previous inequalities imply

$$\begin{aligned} &\frac{\beta + \lambda\alpha}{1+\lambda} Lv \succeq Lw \\ &\frac{\mu\alpha + \beta}{1+\mu} Lv \preceq Lw. \end{aligned}$$

Accordingly,

$$\begin{aligned} \Theta_2(Lv,\, Lw) &\leq \ln \frac{(\beta+\lambda\alpha)(1+\mu)}{(1+\lambda)(\mu\alpha+\beta)} = \ln \frac{e^{\Theta_1(v,\, w)} + \lambda}{e^{\Theta_1(v,\, w)} + \mu} - \ln \frac{1+\lambda}{1+\mu} \\ &= \int_0^{\Theta_1(v,\, w)} \frac{(\mu-\lambda)e^{\xi}}{(e^{\xi}+\lambda)(e^{\xi}+\mu)} d\xi \leq \Theta_1(v,\, w) \frac{1 - \frac{\lambda}{\mu}}{\left(1 + \sqrt{\frac{\lambda}{\mu}}\right)^2} \\ &\leq \tanh\left(\frac{\Delta}{4}\right) \Theta_1(v,\, w). \end{aligned}$$

□

Remark 4.3.3 *In general, it suffices to know $L(\mathcal{C}_1) \subset \mathcal{C}_2$ in order to conclude $\Theta_2(Lv, Lw) \leq \Theta_1(v, w)$. However, a strict contraction depends on the diameter of the image being finite.*[15]

In particular, if an operator maps a convex cone strictly inside itself (in the sense that the diameter of the image is finite), then it is a contraction in the Hilbert metric. This implies the existence of a "positive" eigenfunction (provided the cone is complete with respect to the Hilbert metric), and, with some additional work, the existence of a gap in the spectrum of L (see [10] or [27, Appendix D] for details).

Usually the space $\mathbb{V}$ comes endowed with its own metric, in such a case it is natural to wonder about the strength of the Hilbert metric compared to other metrics. While, in general, the answer depends on the cone, it is nevertheless possible to state an interesting result.

Definition 4.3.4 *A function $\rho : \mathbb{V} \to \mathbb{R}^+$ is called homogeneous of degree one if for all $\lambda \in \mathbb{R}$ and $v \in \mathbb{V}$*

$$\rho(\lambda v) = |\lambda|\rho(v).$$

Remark 4.3.5 *Note that a norm or a linear functional are both homogeneous function of degree one.*

Definition 4.3.6 *A homogeneous function of degree one is called* adapted to a cone $\mathcal{C}$ *if, for each $v, w \in \mathbb{V}$,*

$$-v \preceq w \preceq v \Longrightarrow \rho(v) \geq \rho(w),$$

and $v \in int\, \mathcal{C}$ implies $\rho(v) > 0$.

Lemma 4.3.7 *Let ρ_i be two homogeneous functions of degree one adapted to the cone $\mathcal{C} \subset \mathbb{V}$. Then, given $v,\ w \in int\, \mathcal{C} \subset \mathbb{V}$ for which $\rho_1(v) = \rho_1(w)$,*

$$\rho_2(v - w) \leq \left(e^{\Theta(v,\, w)} - 1\right) \min\{\rho_2(v), \rho_2(w)\}.$$

Proof. We know that $\Theta(v, w) = \ln \frac{\beta}{\alpha}$, where $\alpha v \preceq w \preceq \beta v$. This implies that $-w \preceq 0 \preceq \alpha v \preceq w$, i.e. $\rho_1(w) \geq \alpha\rho_1(v)$, or $\alpha \leq 1$. In the same manner it follows that $\beta \geq 1$. Hence,

$$\begin{aligned}
w - v &\preceq (\beta - 1)v \preceq (\beta - \alpha)v \\
w - v &\succeq (\alpha - 1)v \succeq -(\beta - \alpha)v \\
w - v &\preceq (1 - \beta^{-1})w \preceq (\alpha^{-1} - \beta^{-1})w \\
w - v &\succeq (1 - \alpha^{-1})w \succeq -(\alpha^{-1} - \beta^{-1})w
\end{aligned}$$

which implies

$$\begin{aligned}
\|w - v\| &\leq (\beta - \alpha)\|v\| \leq \frac{\beta - \alpha}{\alpha}\|v\| = \left(e^{\Theta(v,\, w)} - 1\right)\|v\| \\
\|w - v\| &\leq (\alpha^{-1} - \beta^{-1})\|w\| \leq (\frac{\beta}{\alpha} - 1)\beta^{-1}\|w\| \leq \left(e^{\Theta(v,\, w)} - 1\right)\|w\|.
\end{aligned}$$

□

[15] In the theory of Markov processes this corresponds to the so called *positivity improving* (see also Example 4.3.1).

Many normed vector lattices satisfy the hypothesis of Lemma 1.3 (e.g. Banach lattices[16]).

4.3.1 Examples

Perron-Frobenius Theorem

Consider a matrix $L : \mathbb{R}^n \to \mathbb{R}^n$ of all strictly positive elements: $L_{ij} \geq \gamma > 0$. The Perron-Frobenius theorem states that there exists a unique eigenvector v^+ such that $v_i^+ > 0$, in addition the corresponding eigenvalue λ is simple, maximal and positive. There quite a few proofs of this theorem a possible one is based on Birkhoff theorem. Consider the cone $\mathcal{C}^+ = \{v \in \mathbb{R}^2 \mid v_i \geq 0\}$, then obviously $L\mathcal{C}^+ \subset \mathcal{C}^+$. Moreover an explicit computation (see Problem 4.??) shows that

$$\Theta(v, w) = \sum_{ij} \ln \frac{v_i w_j}{v_j w_i}.$$

Then, setting $M = \max_{ij} L_{ij}$, it follows that

$$\Theta(Lv, Lw) \leq 2 \ln \frac{M}{\gamma} := \Delta < \infty.$$

We have then a contraction in the Hilbert metric and the result follows from usual fixed points theorems. Note that, since $\Theta(v, \lambda v) = 0$, for all $\lambda \in \mathbb{R}^+$, the fixed point $v_+ \in \mathbb{R}^n$ is only projective, that is $Lv_+ = \lambda v_+$ for some $\lambda \in \mathbb{R}$; in other words, we have an eigenvalue.

Remark that L^* satisfies the same conditions as L, thus there exists $w^+ \in \mathcal{C}^+$, $\mu \in \mathbb{R}^+$, such that $L^* w^+ = \mu w^+$. Next, define $\rho_1(v) = |\langle w^+, v\rangle|$ and $\rho_2(v) = \|v\|$. It is easy to check that they are two homogeneous forms of degree one adapted to the cone.

In addition, if $\rho_1(v) = \rho_2(v)$, then $\rho_1(L^n v) = \rho_1(L^n w)$. Hence, by Lemma 4.3.7

$$\begin{aligned} \|L^n v - L^n w\| &\leq \left(e^{\Theta(L^n v, L^n w)} - 1\right) \min\{\|L^n v\|, \|L^n w\|\} && (4.3.12) \\ &\leq K\Lambda^n \min\{\|L^n v\|, \|L^n w\|\}, && (4.3.13) \end{aligned}$$

for some constant K depending only on v, w. The estimate 4.3.12 means that all the vectors in the cone grow at the same rate. In fact, for all $v \in \text{int}\mathcal{C}$,

$$\|\lambda^{-n} L^n v - \lambda^{-n} L^n w\| \leq K\Lambda^n.$$

Hence, $\lim_{n\to\infty} \lambda^{-n} L^n v = v_+$.

Finally, consider $\mathbb{V}_1 = \{v \in \mathbb{V} \mid \langle w^+, v\rangle = 0\}$. Clearly $L\mathbb{V}_1 \subset \mathbb{V}_1$ and $\mathbb{V}_1 \oplus \text{span}\{v_+\} = \mathbb{V}$. Let $w \in \mathbb{V}_1$, clearly there exists $\alpha \in \mathbb{R}^+$ such that $\alpha v_+ + w \in \mathcal{C}$,[17] thus

$$\|L^n w\| \leq \|L^n(\alpha v_+ + w) - \alpha L^n v_+\| \leq L\Lambda^n \lambda^n.$$

This immediately implies that L restricted to the subspace $\mathbb{V}_1$ has spectral radius less that $\lambda\Lambda$. In other words, λ is the maximal eigenvalue, it is simple and any other eigenvalue must be smaller than $\lambda\Lambda$. We have thus obtained an estimate of the spectral gap between the first and the second eigenvalue.

[16] A Banach lattice $\mathbb{V}$ is a vector lattice equipped with a norm satisfying the property $\| \, |v| \, \| = \|v\|$ for each $f \in \mathbb{V}$, where $|v|$ is the least upper bound of v and $-v$. For this definition to make sense it is necessary to require that $\mathbb{V}$ is "directed," i.e. any two elements have an upper bound.

[17] this is a special case of the general fat that any vector can be written as the linear combination of two vectors belonging to the cone.

Cones and Q-forms

Here we would like to consider only half of the cone defined by the Q-form in order to apply the present theory. If $L_{ij} > 0$, then we can choose the first quadrant; on the other hand , if $L_{ij} < 0$, then L maps the first into the third quadrant. In both cases a monotone matrix L maps a one sided cone into a one sided cone cone. Here we will consider only the first case and leave the other–essentially identical–to the reader.[18] Consequently, L is a monotone matrix with respect to the standard sector $\mathcal{C}_+$ and $L_{ij} > 0$, then $L\mathcal{C}^+ \subset \mathcal{C}^+$ where, as in the previous example, $\mathcal{C}^+ = \{v \in \mathbb{R}^2 \mid v_i \leq 0\}$. Thus all the results of the previous example apply.

In particular, we have seen that, if $v = (1, \alpha)$, $w = (1, \beta) \in \mathcal{C}_+$, then

$$\Theta(v, w) = \left| \ln \frac{\alpha}{\beta} \right|.$$

Another interesting formula is

$$2 \sinh(\frac{1}{2}\Theta(v, w)) = \frac{|\omega(v, w)|}{\sqrt{Q(v)Q(w)}}. \tag{4.3.14}$$

This means that here exists a relation between the Hilbert metric and the Q-form. To understand this relation better, let us compute

$$\operatorname{diam}(L\mathcal{C}^+) = \sup_{\alpha,\beta>0} \left| \ln \frac{(a + \alpha b)(c + \beta d)}{(a + \beta b)(c + \alpha d)} \right|.$$

Since

$$\frac{a}{c} = \frac{b}{d}(\frac{1 + bc}{bc}) > \frac{b}{d}$$

it follows $\frac{a}{c} \geq \frac{b}{d}$. Thus

$$\operatorname{diam}(L\mathcal{C}_+) = \left| \ln \frac{ad}{cb} \right| = \ln \frac{1 + bc}{bc}, \tag{4.3.15}$$

which implies that, if L is strictly monotone, then $\operatorname{diam}(L\mathcal{C}^+) < \infty$. Accordingly, the rate of contraction of the Hilbert metric is given by

$$\Lambda = \frac{1 - \sqrt{\frac{bc}{1+bc}}}{1 + \sqrt{\frac{bc}{1+bc}}} = \frac{1}{(\sqrt{1 + bc} + \sqrt{bc})^2} = \frac{1}{\sigma(L)^2}, \tag{4.3.16}$$

where the last equality follows by a straightforward computation.

Remark 4.3.8 *It is not immediately clear how to extend the above considerations to the higher dimensional setting discussed in section 4.2. In fact, to do so it is necessary to introduce a different metric [53] of Caratheodory type [72]. We will not do it here but the reader should be aware that such a generalization it is possible.*

[18] An easy way out is to consider L^2 instead of L.

Expanding maps–uniqueness of the a.c. measure

A remarkable fact of Birkhoff theorem is that it applies to infinite dimensional vector spaces. In Example 1.4.1 we have studied the properties of $\mathcal{L}$. A computation similar to the one done there shows that, given a twice differentiable expanding map of the torus, the cone

$$\mathcal{C}_\alpha = \{h \in \mathcal{C}^{(0)}(\mathbb{T}) \mid h \geq 0; \frac{h(x)}{h(y)} \leq e^{\alpha d(x,y)}\} \tag{4.3.17}$$

is invariant. In fact, if $h \in \mathcal{C}_\alpha$

$$\begin{aligned}\mathcal{L}h(x) &= \sum_{z \in T^{-1}x} |D_zT|^{-1}h(z) \leq \sum_{w \in T^{-1}y} \frac{|D_wT|}{|D_zT|}|D_wT|^{-1}h(w)e^{\alpha d(z,w)} \\ &\leq \sum_{w \in T^{-1}y} |D_wT|^{-1}h(w)e^{(\lambda^{-1}\alpha + C)d(x,y)} = e^{(\lambda^{-1}\alpha+C)d(x,y)}\mathcal{L}h(y).\end{aligned}$$

By choosing α large enough, there exists $\sigma \in (\lambda^{-1}, 1)$ such that $\mathcal{L}\mathcal{C}_\alpha \subset \mathcal{C}_{\sigma\alpha}$.

A direct computation shows that the diameter is finite. Accordingly, we have a contraction in the Hilbert metric. This implies that there exists only one invariant measure μ which is absolutely continuous with respect to Lebesgue ($d\mu = h_* dm$). Moreover, if $\rho_1(f) = |\int f|$ and $\rho_2(f) = \|f\|_\infty$, we have that Lemma 4.3.7 applies whereby showing that $\mathcal{L}h \to h_*$ in the sup norm for all $h \in \mathcal{C}_\alpha$, $\rho_1(h) = \rho_2(h_*)$. By arguments similar to the one employed in 4.3.1 it is possible to see that $\mathcal{L}$, viewed as an operator in $\mathcal{C}^{(1)}(\mathbb{T})$, has a maximal eigenvalue one while all the rest of the spectrum is separate by a gap clearly this implies not only the mixing but provides as well an estimate on the mixing rate for $\mathcal{C}^{(1)}(\mathbb{T})$ functions.

4.4 Cones and invariant distributions

Here we use the machinery developed in the previous section to obtain a constructive proof of the existence of the unstable distribution in a special, but very interesting, case.

Lemma 4.4.1 *Given a smooth Symplectic Dynamical Systems with syngularities (X, T, μ), X a symplectic two dimensional manifold, μ the symplectic volume, if the systems is eventually strictly monotone, then $\{E^u(x)\}$ is almost everywhere well defined. Moreover, if $\mathcal{C}(x)$ is continuous, then $\{E^u(x)\}$ is continuous (where it is defined). In addition, if the cone family is strictly monotone, then $\{E^u(x)\}$ is everywhere defined.*

Proof. Let $\mathcal{C}_n(x) := D_{T^{-n}x}T^n\mathcal{C}(T^{-n}x)$ and $\Delta_n(x) := \operatorname{diam}(\mathcal{C}_n(x))$, then Δ_n is decreasing, thus we can define

$$\Delta_\infty(x) := \lim_{n\to\infty} \Delta_n(x).$$

The key consequence of the results of section 4.3 (in particular Examples 4.3.1–Cones and Q-forms) is

$$\begin{aligned}\Delta_\infty(T^m x) &= \lim_{n\to\infty} \operatorname{diam}(D_{T^{m-n}x}T^n\mathcal{C}(T^{-n+m}x)) \\ &= \lim_{n\to\infty} \operatorname{diam}(D_xT^m D_{T^{-n}x}T^n\mathcal{C}(T^{-n}x)) \\ &\leq \frac{1}{\sigma(D_xT^m)^2}\Delta_\infty(x).\end{aligned}$$

Next, let $\Omega = \{x \in X \mid \Delta_\infty(x) = \infty\}$, we claim that $\mu(\Omega) = 0$. In fact, let $B_m = \{x \in X \mid \sigma(D_x T^m) \geq 2\}$, by eventual strict monotonicity of the cone field and Lemma 4.1.2 follows $\mu(\cup_{m\in\mathbb{N}} B_m) = \mu(X)$. In addition, $B_m \supset B_{m_0}$ for all $m > m_0$. Moreover, if $x \in B_m$, then $\Delta_\infty(T^m x) < \infty$ (see 4.3.16). Thus $T^{-n}\Omega \cap B_m = \emptyset$ for all $n \geq m$, and

$$\mu(\Omega) = \lim_{n\to\infty} \mu(T^{-n}\Omega) \leq \lim_{n\to\infty} \mu(X \setminus \cup_{m\leq n} B_m) = 0.$$

Finally, let $\Omega_L = \{x \in X \mid \frac{L}{2} \leq \Delta_\infty(x) \leq L\}$ and suppose $\mu(\Omega_L) > 0$. Then, there exists $n \in \mathbb{N}$ such that $\mu(\Omega_L \cap B_m) > 0$. Consequently, for almost all $x \in \Omega_L \cap B_m$ there exists a return time $\bar{n}m \in \mathbb{N}$ in the past (that is $T^{-m\bar{n}}x \in \Omega_L \cap B_m$). Accordingly,

$$\frac{L}{2} \leq \Delta_\infty(x) \leq \frac{1}{\sigma(D_x T^m)^2} \Delta_\infty(T^{-m\bar{n}}x) \leq \frac{L}{4},$$

which is a contradiction unless $L = 0$. We have so proven that $\mu(\Omega_0) = \mu(X)$. In other words the cones $\mathcal{C}_\infty = \cap_{n\geq 0} \mathcal{C}_n(x)$ is almost everywhere degenerate since, having zero diameter, it consists of a single direction, such a direction is precisely the unstable direction.

To prove the continuity of the above distribution note that the cone family $\mathcal{C}_n(x)$ is continuous. Let x be such that $\Delta_\infty(x) = 0$, then, for each $\varepsilon > 0$, there exists $m \in \mathbb{N}$ such that $\Delta_m(x) < \frac{\varepsilon}{2}$. Then one can chose δ such that the edges of $\mathcal{C}_m(y)$ vary by an amount less than $\frac{\varepsilon}{2}$ if $d(x, y) < \delta$. The result follows then taking into account that the Hilbert metric bounds the angle and that the unstable distribution is contained in $\mathcal{C}_n$ for each $n \in \mathbb{N}$.

The proof of the last fact is obvious: just a simplification of the above arguments. □

With similar techniques it is also possible to construct the stable and unstable foliations, as we will see in chapter 7.

Let us conclude with an interesting simple fact.

Lemma 4.4.2 *A smooth two-dimensional Symplectic Dynamical System (X, T, μ) is Anosov iff it admits a strictly monotone continuous cone family.*

Proof. By Lemma 4.4.1 it follows that the stable and unstable distribution are continuous. But then, by continuity, there exists $\alpha > 0$ and $\sigma > 1$ such that

$$\begin{aligned} &\alpha\sqrt{Q(v)} \leq \|v\| \leq \alpha^{-1}\sqrt{Q(v)} \quad \forall x, \in X \text{ and } v \in E^u(x) \\ &\sigma(D_x T) \geq \sigma \quad \forall x \in X. \end{aligned}$$

Thus,

$$\|D_x T^n v\| \geq \alpha\sqrt{Q(D_x T^n v)} \geq \alpha\sigma^n \sqrt{Q(v)} \geq \alpha^2 \sigma^n \|v\|.$$

Analogously one can obtain the statement for the stable direction by using the cone family given by the complementary cones (see Problem 4.4).

The proof that an Anosov systems admit a continuous strictly invariant cone family is obvious and it is left to the reader.[19] □

Before continuing in the development of the theory it can be helpful to develop and study some more interesting and totally non-trivial examples. To this end are dedicated the next two chapters.

[19] See Problem 7.4.

Problems

4.1 Show that the hypothesis of Theorem 4.1.1 can be relaxed, in particular it holds for smooth systems with singularities (see section 3.5.1). (Hint: Just follow the proof step by step and notice that nothing substantial need to be changed.)

4.2 Construct a strictly invariant cone family for the irrational translation on $\mathbb{T}^2$ (see Examples 1.1.1) and show that it is not measurable. (Hint: For each trajectory choose a point x. At such a point choose the standard cone $\mathcal{C}_+$, let $\mathcal{C}_n^- = \{(v_1, v_2) \in \mathbb{R}^2 \mid 1 + \frac{1}{n} \leq \frac{v_2}{v_1} \leq 2 + \frac{1}{n}\}$ and $\mathcal{C}_n^+ = \{(v_1, v_2) \in \mathbb{R}^2 \mid -2 - \frac{1}{n} \leq \frac{v_2}{v_1} \leq -1 - \frac{1}{n}\}^c$. Then set $\mathcal{C}(T^n x) = \mathcal{C}_n^+$ and $\mathcal{C}(T^{-n}x) = \mathcal{C}_n^-$. Such a cone family is strictly monotone by construction (since $D_x T = 1$), yet the system has obviously zero Lyapunov exponents. Since all the other hypothesis of Theorem 4.1.1 are satisfied, it follows that the above cone family cannot be measurable.)

4.3 Show that for two dimensional symplectic maps the sum of the Lyapunov exponent is zero (*pairing of the Lyapunov exponents*). (Hint: If $\omega(v, w) = 1$ then $1 = \omega(DT^n v, DT^n w) \sim \|DT^n v\|\|DT^n w\|$.)

4.4 Check that $\inf_{v \in \mathcal{C}_+} \sqrt{\frac{Q(Lv)}{Q(v)}} = \left[\inf_{v \in \mathcal{C}_-} \sqrt{\frac{Q(L^{-1}v)}{Q(v)}}\right]^{-1}$, remember that $\mathcal{C}_- = \overline{(\mathcal{C}_+)^c}$. (Hint: see [53])

4.5 Consider $\mathbb{R}^2$ endowed with the scalar product $\langle v, w\rangle_G := \langle v, Gw\rangle$, where $\langle \cdot, \cdot\rangle$ is the standard scalar product and $G > 0$. Show that there exists a change of coordinates $M : \mathbb{R}^2 \to \mathbb{R}^2$ such that, in the new coordinates $\langle \cdot, \cdot\rangle_G$ becomes the standard scalar product.

4.6 Consider the cone $\mathcal{C}$ defined by the two transversal vectors $v_1, v_2 \in \mathbb{R}^2$. This means that $v \in \mathbb{R}^2$ belongs to the cone iff $v = \alpha v_1 + \beta v_2$ with $\alpha\beta \geq 0$. Show that there is a linear change of coordinates $M : \mathbb{R}^2 \to \mathbb{R}^2$ such that $M\mathcal{C} = \mathcal{C}_+$ and $\det M = 1$.

4.7 Show that, in a two dimensional area preserving systems, if the LE are different from zero then there exists and eventually strictly invariant cone family. (Hint: By Oseledets there exists the unstable distributions, then construct the cones around it.)

4.8 Prove that if M is the two by two matrix

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

with $a, b, c \in \mathbb{Z}$, then $M > 0$ iff $a, c > 0$ and $c > \frac{b^2}{a}$.

4.9 Show that a $2d \times 2d$ matrix L of the form

$$L = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A, B, C, D are $d \times d$ blocs, is symplectic, iff $C^*A = A^*C$, $D^*B = B^*D$ and $A^*D - C^*B = \mathbb{1}$. (Hint: Note that, by introducing the matrix

$$J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}$$

the standard symplectic form 4.2.9 can be written in terms of the usual scalar product as
$$\omega((\xi^1, \xi^2),\, (\eta^1, \eta^2)) = \langle (\xi^1, \xi^2),\, J(\eta^1, \eta^2)\rangle.$$
From this point of view the definition of symplectic matrix 4.2.10 can be written as
$$L^* J L = J.$$
A trivial algebraic computation yields now the result.)

4.10 Prove that is L is symplect then $\det L = 1$. (Hint: The determinant of a matrix is nothing else than the volume of the parallelepiped of sides $(Le_1, \dots, Le_{2d})$ (where $e_1, \dots, e_{2d}$ is the standard orthonormal basis of $\mathbb{R}^{2d}$). On the other hand the volume form can be written has $\wedge^d \omega$ (since that is a $2d$ form with the right normalization and the space of $2d$ forms is one dimensional). Thus $\det L = \wedge^d \omega(Le_1, \dots, Le_{2d}) = \wedge^d \omega(e_1, \dots, e_{2d}) = 1$ where we have used the fact that $\omega(Lv, Lu) = \omega(v, u)$. The reader that wants to appreciate the power of the above geometrical interpretation of the determinant and of the external forms can try to prove the statement by purely algebraic means.)

4.11 Show that all symplectic Q-isometrys L (that is $Q(Lv) = Q(v)$) have the form
$$L = \begin{pmatrix} A & 0 \\ 0 & A^{*-1} \end{pmatrix}.$$
(Hint: Start by considering the vector $(0, u)$, $U \in \mathbb{R}^d$, clearly $Q((0, u)) = 0$ thus $Q(L(0, u)) = 0$ if L is a Q-isometry. But if
$$L = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
it follows $\langle Bu, Du\rangle = 0$ for each $u \in \mathbb{R}^d$, that is $B^* D = 0$. The same argument applied to the vector $(u, 0)$ yields $A^* C = 0$. Accordingly, by symplecticity (see Problem 4.9),
$$\begin{aligned} Q(L(v, u)) =& \langle Au + Bv,\, Cu + Dv\rangle = \langle u, (A^* D + C^* B) v\rangle \\ =& \langle u, (\mathbb{1} + 2C^* B) v\rangle \end{aligned}$$
thus $Q(L(v, u)) = Q(v, u)$ iff $C^* B = 0$ which implies $A^* D = \mathbb{1}$.)

4.12 show that if the matrix
$$L = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
is symplectic then
$$L^{-1} = \begin{pmatrix} D^* & -B^* \\ -C^* & A^* \end{pmatrix}$$
(Hint: multiply and use Problem 4.9.)

4.13 Show that the symplectic matrices form a multiplicative group. (Hint: Use the definition and the above problems.)

4.14 A symplectic map L is a Q-isometry iff $L\mathcal{C} = \mathcal{C}$. (Hint: One direction is trivial. On the other hand, if $L\mathcal{C} = \mathcal{C}$ it follows that L maps the boundary, of $\mathcal{C}$, to the boundary. Accordingly, if $\langle v, u\rangle = 0$ it must be

$$0 = \langle Av + bu,\, Cv + Du\rangle. \tag{4.4.18}$$

Choosing in 4.4.18 $u = 0$ yields $A^*C = 0$, choosing $v = 0$ shows that it must be $B^*D = 0$. Thus 4.4.18 yields

$$0 = \langle u,\, (A^*D + C^*B)v\rangle = 2\langle u,\, C^*Bv\rangle.$$

The above equality shows that C^*Bv is parallel to v for each $v \in R^d$, that is $C^*B = \alpha\mathbb{1}$ for some $\alpha \in \mathbb{R}$. If $\alpha = 0$, then $A^*D = \mathbb{1}$ and thus $C = 0$ which is the wanted result. If $\alpha \neq 0$, then B is invertible and $C = \alpha B^{*1}$. But this implies $A = 0$ and hence $-\mathbb{1} = C^*B = \alpha\mathbb{1}$, that is $\alpha = -1$. Accordingly the matrix would have the form

$$L = \begin{pmatrix} 0 & B \\ -B^{*-1} & 0 \end{pmatrix}$$

which sends $\mathcal{C}$ in its complement, contrary to our requirement.)

4.15 Show that a strictly monotone symplectic matrix can be put into the form

$$\begin{pmatrix} \mathbb{1} & \mathbb{1} \\ M & \mathbb{1} + M \end{pmatrix}$$

by multiplying it by Q-isometries on the left and on the right.

4.16 Show that all the Lagrangian subspaces transversal to $V = \{(0,\eta) \in \mathbb{R}^{2d} \mid \eta \in \mathbb{R}^d\}$ can be represented as $\{(\xi,\, U\xi) \in \mathbb{R}^{2d} \mid \xi \in \mathbb{R}^d\}$ for some symmetric matrix U. (Hint: Let $V_U := \{(\xi,\, U\xi) \in \mathbb{R}^{2d} \mid \xi \in \mathbb{R}^d\}$, then $\omega((\xi,\, U\xi),\, (\zeta,\, U\zeta)) = 0$, thus V_U is Lagrangian. On the other hand, if $\tilde{V}$ is Lagrangian, then it is a d dimensional space. Let $\{(\xi_i, \eta_i)\}_{i=1}^d$ be a base for $\tilde{V}$, then $\xi_i \neq 0$ by the transversality assumption and we can define the matrix U via $U\xi := \eta_i$. It is immediate that $\tilde{V}$ Lagrangian implies $U = U^*$.)

4.17 Show that $V_U := \{(\xi,\, U\xi) \in \mathbb{R}^{2d} \mid \xi \in \mathbb{R}^d\}$, $U = U^*$, belongs to the standard cone iff $U \geq 0$.

4.18 Show that given any two transversal lagrangian subspaces V_1, V_2,[20] there exists a symplectic map L such that $LV_1 = \{(\xi, 0)\}$ and $LV_2 = \{(0, \eta)\}$. (Hint: choose coordinates in which V_i are transversal to $V = \{(0,\eta) \in \mathbb{R}^{2d} \mid \eta \in \mathbb{R}^d\}$, then by Problem 4.16 we can write $V_i = \{(\xi, U_i\xi)\}$. Note that, since V_1 and V_2 are transversal, $U_1 - U_2$ must be invertible. The, e.g., set $D = \mathbb{1}$ and $B = (U_1 - U_2)^{-1}$ and check the algebra recalling Problem 4.9.)

4.19 Find a symplectic change of coordinates that transforms the standard form Q into the form Q_h defined by:

$$Q_h((x, y)) = \frac{1}{2}(\langle x, x\rangle - \langle y, y\rangle),$$

[20] Recall that two space are transversal iff $V_1 \cap V_2 = \emptyset$.

and draw the associate cone. (Hint: Consider

$$\begin{aligned} x &= \tfrac{x'-y'}{\sqrt{2}} \\ y &= \tfrac{x'+y'}{\sqrt{2}}.) \end{aligned}$$

4.20 Prove that $(\mathbb{V}, \preceq)$ is a lattice. (Hint: The convexity implies the transitivity of the relation. The other properties can be checked directly.)

4.21 Let $C \in \mathbb{R}^n$ be a strictly convex compact set. For each two point $x, y \in C$ consider the line $\ell = \{\lambda x + (1 - \lambda y) \mid \lambda \in \mathbb{R}\}$ passing through x and y. Let $\{u, v\} = C \cap \ell$ and define

$$\Theta(x, y) = \left| \ln \frac{\|x-u\|\|y-v\|}{\|x-v\|\|y-v\|} \right|$$

(the logarithm of the cross ratio). Show that Θ is a metric in C (the Hilbert metric). Show that the distance of $x \in \text{int}\, C$ from ∂C is infinite. (Hint: The only non trivial task is to check the triangle inequality. Consider three points $x, y, z \in C$. If the points are collinear then the proof is easy. If they are not the consider the plane defined by them, we have now a two dimensional problem, thus it suffices to prove the result in $\mathbb{R}^2$. Consider the Figure 21 and remember that the cross ratio

$$R(x, y, u, v) = \frac{\|x-u\|\|y-v\|}{\|x-v\|\|y-v\|}$$

is a projective invariant. Then

$$\begin{aligned} R(x, z, u, v) &= R(x, w, x', y') \geq R(x, w, \alpha, \beta) \\ R(y, z, a, b) &= R(w, y, x', y') \geq R(w, y, \alpha, \beta) \end{aligned}$$

and the result follows since $R(x, w, \alpha, \beta)R(w, y, \alpha, \beta) = R(x, y, \alpha, \beta)$.)

4.22 Prove the same as the previous Problem for convex sets in a projective space.

4.23 Check that the metric defined in Problem 4.21 and Problem 4.22 applied to the projectivization of convex cones yields to the same metric defined in (4.3.11).(Hint: Let $\mathcal{C} \subset \mathbb{V}$ be a convex cone. The projectivization consists in considering the space $\mathbb{P}$ of the equivalence classes $[v]$ with respect to the equivalence relation $v \sim w$ iff $v = \lambda w$ for some $\lambda \in \mathbb{R}^+$. Let $\tilde{\mathcal{C}} = \{[v] \in \mathbb{P} \mid v \in \mathcal{C}\}$. Clearly $\tilde{C}$ is convex in the projective space $\mathbb{P}$.[21] So, if $[x], [y] \in \tilde{\mathcal{C}}$, the Hilbert metric is defined by the points $[u], [v] \in \partial\tilde{\mathcal{C}}$ on the line determined by $[x]$, $[y]$. By normalizing properly one obtains $[u] = [-\alpha x + y]$ and $[v] = [\beta x - y]$ form which the result follows.)

4.24 Hilbert metric for a disc and the half plane–hyperbolic geometry.

4.25 Kobayashi and Caratheodory metrics....

4.26 Find an explicit formula for the Hilbert metric for the cone $\mathcal{C}_\alpha$ defined in (4.3.17).

[21] The lines in $\mathbb{P}$ are given by $[\alpha v + \beta w]$ where $\alpha, \beta \in \mathbb{R}$ where $[v] \neq [w]$.

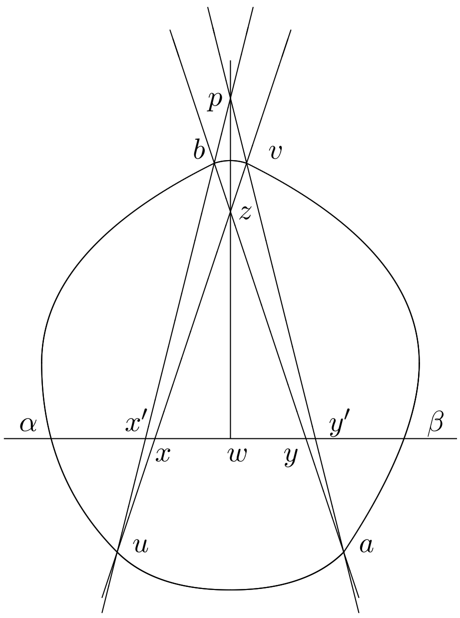

Figure 4.1: Hilbert metric

4.27 Show that the Perron-Frobenius operator associated to a smooth expanding map of the circle has a spectral gap as an operator on $Lip(\mathbb{T}^2)$. (Hint: Check that there exists $b \in \mathbb{R}^+$ such that the norm

$$\|h\| := \|h\|_\infty + b\|h\|_{Lip}$$

is adapted to the cone. Define $\mathbb{V} = \{h \in Lip(\mathbb{T}^2) \mid \int h = 0\}$, notice that $\mathcal{L}\mathbb{V} = \mathbb{V}$. Then, for each $h \in \mathbb{V}$ there exists$\rho \in \mathbb{R}^+$ such that $h + \rho h_* \in \mathcal{C}_\alpha$, so

$$\|\mathcal{L}^n h\| = \|\mathcal{L}^n(h + \rho h_*) - \rho h_*\| \leq K\Lambda^n \rho.$$

Thus the spectral radius of $\mathcal{L}|_\mathbb{V}$ is less than Λ.)

4.28 Estimate the rate of mixing for Lipschitz functions for a smooth expanding map of the circle (Hint: use the spectral gap of the previous Problem.)

4.29 Prove that any continuous fraction of the form

$$\cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}}$$

$a_i > 0$ is convergent provided the series $\sum_{n=1}^\infty a_n$ is divergent. (Hint: Let

$$\prod_{i=1}^n \begin{pmatrix} 1 & a_{2(n-i)} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a_{2(n-1)+1} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ u \end{pmatrix} = \begin{pmatrix} \beta_n \\ \alpha_n \end{pmatrix}$$

and verify, by induction, that $\frac{\alpha_n}{\beta_n}$ is exactly the $2n$ truncation of the continuous fraction. Thus the continuous fraction is a projective coordinate for the vector (α_n, β_n). Consider the cone $\mathcal{C}_+ = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0; y \geq 0\}$. Then, for each $a, b \in \mathbb{R}^+$, holds

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \mathcal{C}_+ \subset \mathcal{C}_+.$$

The result follows by computing the Hilbert metric contraction.

For a different approach see [73][Th14.1].)

4.30 Prove 4.3.16.

Notes

The point of Wojtkowski Theorem is that no estimate on the cone contraction is needed. To appreciate the usefulness of this fact, one can try to prove the existence of the LE via direct estimates for the Levowich map in example 4.1.1.

For generalization of metrics on cones, see [72]

CHAPTER 5

Description of Billiard systems

Billiards are very widely studied model systems.

In general they consists of a material point confined in some region of $\mathbb{R}^n$ or $\mathbb{T}^n$ with piecewise smooth boundaries;[1] in the simplest situation such a point moves with constant velocity until it reaches the boundary, and at the boundary it undergoes an elastic reflection. Such models include, e.g., a system of n hard spheres that interacts via elastic collisions (see section 5.5); the importance of such a systems as a basic model in statistical mechanics can be hardly overestimated.

These systems are conceptually extremely simple, yet they have an unpleasant feature: they lack smoothness. As we will see in the following there are three main type of non-smoothness: a) tangent collisions; b) collision with a corner; c) accumulation of infinitely many collisions in a finite time. Due to such pathologies these models, in spite of their simplicity, may present some incredibly annoying complications in their treatment.

Let $\mathcal{B}$ be the region in which the point is allowed to move and suppose that $\partial\mathcal{B}$ is a finite union of smooth manifolds with boundary. Clearly the motion can be seen as a flow ϕ_t on the unitary tangent bundle of $\mathcal{B}$ (in fact, given the initial position and the initial velocity the following motion is uniquely determined, moreover the modulus of the velocity will be constant through the motion, so it can be assumed equal to one without loss of generality).[2]

It can be checked directly that the flow is symplectic (Hamiltonian) in $\widetilde{X} := \mathcal{B} \times \mathbb{R}^n$ (see problem Problem 5.5). So, calling m the measure induced by Lebesgue on $X \equiv \mathcal{B} \times \{v \in \mathbb{R}^n \mid \|v\| = 1\}$, (X, ϕ_t, m) is a smooth flow with collisions (crf. Examples 3.6.1).

5.0.1 Examples

Polygonal Billiards

The name is self-explanatory: the domain $\mathcal{B}$ is a polygon. The simplest case is probably a rectangle: $\mathcal{B} = [0, a] \times [0, b] \subset \mathbb{R}^2$. Although the notion is fairly trivial, to study it we will employ a neat trick that has many other applications (e.g., see **??**). Consider a trajectory $x+vt$ that reaches the wall $\ell_1 := \{(a, y)\}$. The law of reflection states that, if $v = (v_1, v_2)$, the reflected velocity is $(-v_1, v_2)$. Now define the map $R_a(x, y) = (2a - x, y)$. This is a reflection ($R_a^2 =$ identity) with respect to the wall $\{(a, y)\}$. Remark that $R_a\mathcal{B} = [a, 2a] \times [0, b]$, moreover $DR_a(-v_1, v_2) = v$. This means that, in the reflected box $R_a\mathcal{B}$, the reflected velocity is equal to the velocity before reflection.

[1] Although one can easily consider billiards in a region of a Riemannian manifold with piecewise smooth boundaries, in this case the motion in the interior is just the geodesic flow; see [14] for such a general setting.

[2] A little thought will convince the reader that two motions with initial velocities that differ only in modulus will be exactly the same apart from the fact that they are run at different speeds.

The above algebraic discussion corresponds to a very intuitive geometrical fact: if the wall is a mirror, then the trajectory in the mirror is the continuation of the trajectory before collision (see figure **??**).

After noticing this it is quite clear that one can understand better the trajectory in the "universal covering' of the box obtained be reflecting the box repeatedly with respect to its walls. In this covering the trajectory is simply a straight line and the trajectory in the original box is obtained by undoing the reflections (for the more mathematical inclines let us say that the plane is covered by equal boxes that are identified via reflections, see Problem 5.1). It is then obvious that, given the original velocity v only four velocities are possible: $(\pm v_1, \pm v_2)$. In fact, if we identify the opposite sides in figure **??** (that is, side ℓ_3 with side ℓ_1, side ℓ_2 with $R_a\ell_1$ and so forth) we obtain exactly a flat torus with sides twice as long as teh ones of the original rectangel. In addition, the motion on such a torus corresponds precisely to the flow at unit speed in direction v. in other words the motion is equivalent to the rigid translations (geodesic flow) on the associated torus.

Accordingly, the motion is ergodic only if $\frac{v_1 b}{v_2 a}$ is irrational (see **??**).

Circular Billiards

In this case $\mathcal{B}$ is a disk of radius r. For convenience, let us center it at the origin of a Cartesian coordinate frame. Let us consider a point that has just collided with the boundary at the position $rn(\theta) := r(\cos\theta, \sin\theta)$, where θ is the angle with the x axis counted counterclockwise, and has velocity $v(\theta-\varphi) := (-\sin(\theta-\varphi), \cos(\theta-\varphi))$, which means that the velocity forms an angle φ with the tangent at the collision point. Accordingly, the trajectory will move along the cord of length $2r\sin\varphi$ and collide with the angle $\pi-\varphi$ which, after reflection, will be φ again.

This phenomena is nothing else than the conservation of the angular momentum (for the mechanical inclined) or of the Claroit integral (for the differential geometers).

All the above implies that, if $\frac{\varphi}{\pi} \in \mathbb{Q}$, then the motion will be periodic, otherwise the collision point will perform an irrational rotation on the boundary. In fact, let us choose as coordinates the distance τ form the last collision point computed along the trajectory; the distance s, computed along the circumference, of the last collision point from a fixed point on the circumference; and the angle φ. Then the phase space is

$$X = \{(\tau, s, \varphi) \in [0, r] \times S^1 \times [0, \pi] \mid 0 \le \tau \le 2r\sin\varphi\}$$

and the flow is noting else that a suspension flow (see ???) with ceiling function $2r\sin\varphi$ constructed on the map T defined by

$$T(s, \varphi) = (s + \frac{2}{\varphi}, \varphi).$$

At the same time the middle point of the cords between two consecutive collisions will describe an irrational rotation on the circle of radius $r(1-\cos\varphi)$. This last circle is called *caustic*; the name derives from optic because if the trajectory is run by a beam of light that is the place with the highest luminosity.[3] Note that this means that the trajectory under consideration (if $\varphi/\pi \notin \mathcal{Q}$) covers densely a two dimensional torus in the three dimensional space and it is ergodic restricted to it.

[3] In ancient Greek caustic ($\kappa\alpha\upsilon\sigma\tau\iota\kappa\acute{o}\zeta$) means "that burns". Of course, that would be an important concept if you want, e.g., burn a Roman ship. (check the Optic of Euclid)

The above examples correspond to very regular motions ("integrable motion") that is exactly the opposite of what we mean to investigate. Unfortunately, to progress in the direction we are interested in many more technical tools are needed. Yet, before going on with general facts and definitions let us anticipate two concrete examples that will be particularly relevant in the following.

5.1 Sinai Billiard

The simplest example of Sinai billiards (introduced in [62] and studied in [63]) are given when $\mathcal{B} \subset \mathbb{T}^2$. More precisely, given a disk D, centered at the origin and with diameter $r < \frac{1}{2}$, let $\mathcal{B} = \mathbb{T}^2 \backslash D$. Calling $(x, v) \in \mathcal{B} \times \mathbb{R}^2$ the position and the velocity, respectively, the motion is described by a free flow

$$\phi_t(x, y) = (x + vt, v), \tag{5.1.1}$$

provided $\|x + vt\| \geq r$, that is provided the motion does not exist $\mathcal{B}$. When $x \in \partial\mathcal{B} = \partial D$ a collision takes place. Of course, at the collision it must be $\langle x, v\rangle \leq 0$, the velocity points toward D, otherwise the point would not have reached the obstacle D but rather would be flowing away from it. The collision law is, as already said, an elastic collision–namely the total energy and the momentum tangential to the collision plane must be preserved. Thus, calling v_- the velocity before collision and v_+ the velocity after collision, we require

$$\|v_+\| = \|v_-\| \ ; \quad \langle Jx, v_-\rangle = \langle Jx, v_+\rangle,$$

where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

so that $\langle Jx, x\rangle = 0$, that is $r^{-1}Jx$ is a unit vector tangent to the disk and oriented counterclockwise. This implies:

$$v_+ = v_- - \frac{2}{r^2}\langle x, v\rangle x. \tag{5.1.2}$$

5.1.1 Flow

From the above discussion it is clear that (X, ϕ^t, m) is a smooth flow with collisions, the only property that need to be checked is (3.6.14).

Let us call $V(x, v) = (v, 0)$ the vector field generating ϕ^t. A useful fact is the following.

Lemma 5.1.1 *If $w \in \mathcal{T}_\xi X$ and $\langle w, V(\xi)\rangle = 0$, then $\langle d\phi^t w, V(\phi^t(\xi))\rangle = 0$.*

PROOF. If no collision takes place, then the statement it is obvious by equation (5.1.1) and since for each $w = (w_1, w_2) \in \mathcal{T}X$ it must be $\langle w_1, v\rangle = 0$ (just differentiate $\|v\|^2 = 1$). Let us see what happens at collision.

Given the tangent vector $w = (w_1, w_2)$ at the point $\xi \in X$, we can consider the curve $\gamma(s) = \xi + ws$ that generates it ($\gamma'(0) = w$). Suppose that the next collision takes place with an angle φ. If we refer to the Figure **??** all we need to compute is the relation between h and l. A bit of geometry shows that

$$h = s \arctan\varphi + \mathcal{O}(s^2); \quad l = \frac{s}{\cos\theta} \arctan\varphi + \mathcal{O}(s^2) = s \arctan\varphi + \mathcal{O}(s^2).$$

Thus, if τ is the collision time of the trajectory starting at ξ and $\tilde{\gamma}(s) = \phi^{\tau_+}(\gamma(s))$, we have $\tilde{\gamma}'(0) = d_\xi \phi^{\tau_+} w := \tilde{w}$, and, calling v_+ the velocity after reflection, $\langle v_+, \tilde{w} \rangle = 0$, which proves the lemma. □

This means that in this case there is a particularly simple way to quotient out the flow direction: consider only vectors perpendicular to the flow.

5.1.2 Poincaré map

For many purposes it is useful to view the Sinai billiards as a symplectic map from a two dimensional domain to itself. Such a reduction is obtained via a general technique widely used in dynamical system: a Poincaré section (see 1.2). A Poincaré section consists in introducing some codimension one manifolds in the phase space X and then defining a map from such manifolds to themselves in such a way that to each point is associated its first return to the manifolds (if it exists). Let us be more concrete.

Historically the choice of the section to realize a Poincaré map as been based on $\partial\mathcal{B}$. In our case this consists of the boundary of the disk, that is a circle. Of course, it is also necessary to specify the velocity. Clearly there are two possibilities: one can consider velocities just before collision, which means $\langle x, v \rangle \leq 0$, (this is the Poincaré map from before collision to before collision) or one can consider the velocity just after collision, meaning $\langle x, v \rangle \geq 0$, (that is the Poincaré map from just after collision to just after the next collision). The two choices are equivalent, let us make the second.

If we define the velocity by the angle φ between v and the tangent (directed clockwise) to the disk at the collision point, then the phase space is $\mathcal{M} = S^1 \times [0, \pi]$.

We can then define a map $T : \mathcal{M} \to \mathcal{M}$ in the following way: for each $\xi \in \mathcal{M}$, let $T\xi$ be the point just after the next reflection (if such a reflection exists). Note that, if no reflections would occur, almost all the trajectories would fill $\mathbb{T}^2$ densely,[4] hence T is defined almost everywhere.

It is natural to use as coordinate on the boundary the distance s, computed counter-clockwise along the circle, from a given point. If we want to compute the induced invariant measure on the Poincaré section $\mathcal{M}$, according to section ?, we have to introduce the change of coordinates $\Xi : \mathcal{M} \times [0, \delta] \to X$ defined by

$$\Xi(s, \varphi, t) = (rn(sr^{-1}) + v(sr^{-1} + \varphi)t, sr^{-1} + \varphi - \frac{\pi}{2}),$$

where $n(\theta) = (\cos\theta, \sin\theta)$, $v(\theta) = (\sin\theta, -\cos\theta)$. In this coordinates a point is determined by its collision data (s, φ) and the time t past from the last collision.

A direct computation shows that

$$\begin{aligned}
\det \Xi &= \begin{vmatrix} -v(sr^{-1}) + r^{-1}n(sr^{-1} + \varphi)t & n(sr^{-1} + \varphi) & v(sr^{-1} + \varphi) \\ r^{-1} & 1 & 0 \end{vmatrix} \\
&= \begin{vmatrix} -v(sr^{-1}) & n(sr^{-1} + \varphi) & v(sr^{-1} + \varphi) \\ 0 & 1 & 0 \end{vmatrix} \\
&= \begin{vmatrix} -v(sr^{-1}) & v(sr^{-1} + \varphi) \end{vmatrix} = \sin\varphi.
\end{aligned}$$

[4] Since for almost all velocities v we would have an irrational translation on $\mathbb{T}^2$.

So, given a set $A \subset \mathcal{M}$, calling $A_\varepsilon = \cup_{t=0}^{\varepsilon} \phi^t A$,

$$\mu(A) := \frac{1}{\varepsilon} m(A_\varepsilon) = \frac{1}{\varepsilon} \int_{A_\varepsilon} |det(\Xi)| ds d\varphi dt = \int_A \sin\varphi ds d\varphi.$$

Thus $d\mu = \sin\varphi ds d\varphi$ and $(\mathcal{M}, T, \mu)$ is a Dynamical Systems.

It is interesting to notice that μ becomes degenerate for $\varphi \in \{0, \pi\}$, which correspond to tangent collisions. Another annoying feature of the above choice is that some trajectories never meet the boundary of the disk (for example consider the initial condition $x = (1, 0)$, $v = (0, 1)$) and other will travel an arbitrarily long time before the next collision.[5] These facts, although not catastrophic, may look unpleasant to someone. It is therefore relevant to notice that there are several other possible choices for the Poincaré section, each one with is own advantages and disadvantages. Let us see a couple of them.

Consider the fundamental domain $Q = [-\frac{1}{2}, \frac{1}{2}]^2$ of $\mathbb{T}^2$, choose ∂Q as the basis of the Poincaré section. Of course ∂Q it is not a smooth manifold (it consists of four lines). This problem is easily solved by extending the concept of Poincaré section to the case in which the section Σ is a finite (or even countable) union of smooth manifolds; a quick look at section ?? will convince the reader that this generalization is indeed immediate. This section has the the advantage of a simple structure, that there is a maximal time from Σ to itself, yet it does not solves the problem of the degeneration of the measure. Here as well we have problems with the trajectories that meet the section at very small angles.

To overcome such a problem one can choose the section $\Sigma \times [\delta, \pi - \delta]$. It is easy to see that if δ is chosen small enough then the only effect is to skip at most one crossing of the boundary Σ.

We will keep using the relation between the two dynamical systems (X, m, ϕ_t) and $(\mathcal{M}, \mu, T)$. In particular it is convenient to define the cone family on all $\mathcal{T}X$ instead that only on $\mathcal{T}\mathcal{M}$. We will see that an invariant cone family on $\mathcal{T}X$ induces an invariant cone family on $\mathcal{T}\mathcal{M}$.

5.1.3 Singularity manifolds

In this subsection we will study more precisely the singularities of the system and we will verify that they belong to the general setting developed in 3.5.1. We will consider two different Poincaré section to give the reader a more complete idea of the situation.

Let us start with the classical section *just after collision*. As already mentioned the phase space is $\mathcal{M} = S^1 \times [0, \pi]$. Clearly the only singularities of the map correspond to coordinates where the next collision is a tangent one. To analyze such a pathology it is more convenient to look at the billiard on the universal covering of the torus. In such a covering the obstacles will form a a lattice and the particles moves along a straight line between collisions (see figure **??**).[6]

The particle with coordinates (s, φ), just after collision, will move in the direction $v(r^{-1}s + \varphi)$ with unit speed.[7] Hence, if $C \in \mathbb{R}^2$ is the coordinate of the center of the obstcle with

[5] This property is called *infinite horizon*. We will discuss it further in the sequel.

[6] This trick is very similar to the one used at the beginning of the chapter to discuss rectangular billiards, only now we take advantage of the periodicity of the torus rather than the invariant properties of the domain with respect to the reflections.

[7] Remember the convention $n(\theta) := (\cos\theta, \sin\theta)$ and $v(\theta) := (\sin\theta, -\cos\theta)$.

which the next collsion will take place, the condition for a tangent collision reads

$$rn(r^{-1}s) + tv(r^{-1}s + \varphi) = C \pm rn(r^{-1}s + \varphi). \tag{5.1.3}$$

Where $t = t(s, \varphi)$ is the collision time. From (5.1.3 we can imediately extract two interesting informations multipling it by $n(r^{-1}s + \varphi)$ and $v(r^{-1}s + \varphi)$ respectively

$$\begin{aligned}
&\langle C, v(r^{-1}s + \varphi)\rangle = t + r\sin\varphi > 0\\
&F(s, \varphi) := \langle C, n(r^{-1}s + \varphi)\rangle - r\cos\varphi \pm r = 0
\end{aligned}$$

Taking the derivative of F with respect to φ we get

$$-r\sin\varphi + \langle C, v(r^{-1}s + \varphi)\rangle = t > 0,$$

thus we can apply the implicit function theorem and conclude that the manifold corresponding to this discontinuity can be represented as the graph of a function $\varphi(s)$. In addition,

$$\frac{d\varphi}{ds} = -(\frac{1}{r} + \frac{\sin\varphi}{t}) < 0. \tag{5.1.4}$$

Since there are infinitely many obstacles with which the next collision can take place, there must be countably many discontinuity line (some of them are schematically represented in figure **??**)

To analize the preimages of the boundary of the section one can proced in analogy with what we have done before, equation (5.1.3) in this case beomes

$$rn(r^{-1}s) + tv(r^{-1}s + \varphi) = C \pm rn(r^{-1}s + \varphi \pm \delta). \tag{5.1.5}$$

From (5.1.5) we obtain

$$\frac{d\varphi}{ds} = -(\frac{1}{r} + \frac{\sin\varphi}{t + r\sin\delta}) < 0. \tag{5.1.6}$$

Clearly the map is smooth up to, and including, this type of discontinuity, not so for the tangencies. In fact, it is easy to verify that the map is continuous crossign a tanency line (see Problem 7.7) but we will see imediately that it is not differentiable. By the discussion of section 5.3 (see in particular fromulae (5.3.7) and (5.3.8)) it follows that if the next collision takes place with an angle $\varphi \notin [\delta, \pi - \delta]$, then calling τ_1 the time up to the tangent collision and τ_2 the time from the tangent collision to the next, we have the formula

$$DT = \begin{pmatrix} 1 + \frac{2\tau_1}{r\sin\varphi} & \frac{2}{r\sin\varphi} \\ \tau_2(1 + \frac{2\tau_1}{r\sin\varphi}) & 1 + \frac{2\tau_2}{r\sin\varphi} \end{pmatrix}.$$

Clearly the norm of DT is bounded by a constant times $\frac{1}{\sin\varphi}$ (if in doubt do Problem 7.8). Now, if a point has distance ε from the singularity line, it will land at a distance $\sqrt{\varepsilon}$ from the tangency, which meand that there exists a constant $c_t > 0$ such that, calling $\mathcal{S}$ the syngularity line and ξ the point

$$|\sin\varphi| \geq c_t\sqrt{d(\xi, \mathcal{S})}.$$

Which means that the Derivative blows up only as a square root getting close to the singularity. By similar considerations it is possible to veryfy also that the second derivative blows up polinomially (see Problem 7.9).

5.2 Bunimovich Stadium

These billiards have been introduced [12] and first studied [13] by Bunimovich. In this case the table of the billiard is a convex subset of $\mathbb{R}^2$. The simplest, and original, one consists in two half circles joined by two straight lines (see Figure **??**).

The name "stadium" is due to the shape of the domain $\mathcal{B}$ in which the motion takes place. The only difference is that now the curvature of the boundary is either zero (collisions against the straight segments) or negative (collision against the half circles).

5.2.1 Flow

We have seen that the flow in a square or in a circle is well defined and rather regular. Clearly the only relevant discontinuity in the Bunimovich Billiard arise when a trajectory hit the joining between the circumference and the straight lines.

5.3 Collision map and Jacobi fields

To compute, in general, the collision map it is helpful to introduce appropriate coordinates in $\mathcal{T}X$. A very interesting choice is constitute by the *Jacobi fields*.[8] Let X_- be the set of configurations just before collision. For each $(x, v) \in X \backslash X_-$ there exists $\delta > 0$ such that

$$\phi_t(x, v) = (x + vt, v) \quad 0 \le t \le \delta.$$

Let us consider the curve in $\mathcal{X}$

$$\xi(\varepsilon) = (x(\varepsilon), v(\varepsilon)),$$

with $\xi(0) = (x, v)$ and $\|v(\varepsilon)\| = 1$.

For each t such that $\phi_t(\xi(0)) \notin X_-$, let

$$\xi(\varepsilon, t) = (x(\varepsilon, t), v(\varepsilon, t)) = \phi_t(\xi(\varepsilon)).$$

The Jacobi field $J(t)$ is defined by

$$J(t) \equiv \left.\frac{\partial x}{\partial \varepsilon}\right|_{\varepsilon=0}.$$

Note that, since $x(0, t) \notin X_-$, for $s < \delta$

$$\xi(\varepsilon, t+s) = \xi(\varepsilon, t) + (v(\varepsilon, t)s, 0),$$

so

$$J'(t) = \frac{dJ(t)}{dt} = \left.\frac{\partial v(\varepsilon, t)}{\partial \varepsilon}\right|_{\varepsilon=0}.$$

That is, $(J(t), J'(t)) = d\phi_t \xi'(0)$.

[8]The Jacobi Fields are a widely used instrument in Riemannian geometry (see [31]) and have an important rôle in the study of Geodetic flows, although we will not insists on this aspect at present. Here they appear in a very simple form.

At each point $\xi = (x, v) \in X$ we choose the following base for $\mathcal{T}_\xi X$:[9]

$$\eta_0 = (v,\, 0); \quad \eta_1 = (v^\perp,\, 0); \quad \eta_2 = (0,\, v^\perp);$$

where $\|v^\perp\| = 1$, $\langle v,\, v^\perp\rangle = 0$.

The vector η_0 corresponds to a family of trajectories along the the flow direction and it is clearly invariant; η_1 to a family of parallel trajectories and η_2 to a family of trajectories just after focusing. It is very useful the following graphic representation. We represent a tangent vector by drawing a curve that it is tangent to it. A curve in $\mathcal{T}X$ is given by a base curve that describes the variation of the x coordinate equipped with a direction at each point (specified by an arrow) which show how varies the velocity (see Figure ??).

A direct check shows that each vector η perpendicular to the flow direction will stay so (see Lemma 5.1.1), i.e.

$$\langle d\phi_t \eta,\, (v_t,\, 0)\rangle = \langle d\phi_t \eta,\, d\phi_t(v,\, 0)\rangle = \langle \eta,\, (v,\, 0)\rangle = 0.$$

So the free flow is described by

$$d\phi^t \eta_0 = \eta_0; \quad d\phi^t \eta_1 = \eta_1; \quad d\phi^t \eta_2 = \eta_2 + t\eta_1,$$

that is, in the above coordinates

$$d\phi^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & t & 1 \end{pmatrix}. \tag{5.3.7}$$

Let us see now what happens at a collision.

Let $x_0 \in \partial\mathcal{B}$ be the collision point and let $\xi_c = (x_0, v)$ be the configuration at the collision. We want to compute $R_\varepsilon := d_{\phi^{-\varepsilon}\xi_c}\phi^{2\varepsilon}$, that is the tangent map from just before to just after the collision. Clearly $R_\varepsilon \eta_0 = \eta_0$. From the Figure **??** follows that, if $\gamma(s)$ is the curve associated to η_1 at the point $\phi^{-\varepsilon}\xi_c$,

$$d\phi^{2\varepsilon}\gamma(s) = \left(v_+^\perp \left[s + \varepsilon \frac{2s}{r \sin\varphi} \right],\, \frac{2s}{r\sin\varphi} \right) + \mathcal{O}(s^2)$$

where r is the radius of the osculating circle (that is the circle tangent to the boundary up to second order) which is the inverse of the curvature $K(x_0)$ of the boundary at the collision point.

The above equation means that

$$J(\varepsilon) = (1 + \frac{2\varepsilon K(x_0)}{\sin\varphi}) v_+^\perp.$$

Accordingly, calling $R = \lim_{\varepsilon\to 0} R_\varepsilon$ the collision map, we have

$$R\eta_1 = \eta_1 + \frac{2K}{\sin\varphi}\eta_2; \quad R\eta_2 = \eta_2.$$

[9] Here $v^\perp = Jv$ with

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Hence,

$$DR = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{2K}{\sin\varphi} \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.3.8}$$

The above computations provide the following formula for the derivative of the Poincaré section from the boundary of the obstacle, just after collision, to the boundary of the obstacle just after the next collision

$$DT = \begin{pmatrix} 1 & \frac{2K}{\sin\varphi} \\ \tau & 1 + \frac{2\tau K}{\sin\varphi} \end{pmatrix}, \tag{5.3.9}$$

where τ is the flying time between the two collisions and φ the collision angle.

Formula (5.3.9) is sometimes called *Benettin formula* (e.g., [45]).

5.4 Different Tables, different games

Let us start with a bit of classification.

Definition 5.4.1 *Here are some standard classes of billiards:*

- *dispersing billiards are billiards with the boundary $\partial\mathcal{B}$ of the table is a finite union of stricly convex manifolds with boundary (this are often called Sinai billiards as well)*
- *semi-dispersing billiards are billiards with the boundary $\partial\mathcal{B}$ of the table is a finite union of stricly convex manifolds with boundary*
- *convex billiards are billiards where the tale $\mathcal{B}$ is a convex set.*

The remainder of the section is devoted to a more explicit description of several concrete examples of the above cases.

5.4.1 Dispersing

We have already seen the standard Sinai billiard in section 5.1. In general several convex obstacles may be present and they are not necessarily disjoint (see figure **??**). One main issue in this class of billiards is the distinction between finite and infinite horizon. Finite Horizon means that there is a maximal time after which a collision must take place, infinite horizon means that there exists trajectories that never experience a collision.[10] The relevance of such concept stems from the fact that orbit with no collision have zero Lyapunov exponents, hence the corresponding billiards cannot be uniformly hyperbolic.

Infinite Horizon

As already mentioned we have already seen the prototypical example in this class, yet it may be instructive to analize its properties a bit further. Consider the Sinai billiard described in section 5.1 and let r_1 the radius of the obstacle. Clearly it is necessary $r_1 < 1/2$ to have no self intersections of the obstacle. It is also obvious that if $r_1 < 1/2$ then there are trajecotirs

[10] Note that the other possibility (all the trajectories experience a collision in finite time but there does not exists an upper bound for such a time) cannot take place (see Problem 5.11).

that never collide. Let us study such trajectories a bit more in detail. First of all, since the sistem has a square simmetry, it is enough to consider trajectories with velocity in the first half of the first quadrant, i.e. velocities parallel to the directions $(1, \omega)$ with $\omega \in [0, 1]$.[11]

Let us consider the motion with no obstacle (a traslation on the torus) and see if there are trajectories that never enter in the region $\|(x, y)\| \leq r_1$. Clearly such trajetories are trajectories for the billiard systems as well and precisely the trajecotries that never experince a collision. For such trajectories it is particularly convenient to consider the poincarè section determined by the line $S : \{x = -1/2\}$. If we look at the motion only when the particle intersects such a line we have that the corresponding map is given by $Ty = y + \omega \mod 1$, that is a rotation by ω of the circle $(-1/2, 1/2]$. If $\omega \notin \mathcal{Q}$ then the map T is ergodic (see **??**) and this menas that the trajectory will eventually collide. On the contrary, if $\omega = p/q$, $p, q \in \mathbb{N}$ and with no common divisors, then all the orbit will be periodic of period q (see **??**) and it may be possible that some of them never collide.

Notice that a point in S with velocity parallel to $(1, \omega)$ will experience before a collision before meeting S again only if $y \in [-\omega/2 - r_1\sqrt{1 + \omega^2}, -\omega/2 + r_1\sqrt{1 + \omega^2}]$. On the other hand, since the orbit of the point y has lenght q and because it is restricted to points of the type $y + n/q \mod 1$, which are exactly q, it follows that the orbit consists exactly of all such points. Accordingly, the orbit can avoid only intervals of size less than $1/q$. We can then conclude that there are orbits of the type p/q that never collide if and only if

$$2r_1\sqrt{1 + \frac{p^2}{q^2}} < \frac{1}{q}. \tag{5.4.10}$$

If $q = 1$ then for $p = 0$, we have the already know result $r_1 < 1/2$; for $p = 1$ there can be no collisions only if $r_1 < \frac{1}{2\sqrt{2}}$. For $q \leq 2$ there are always collisions if $r_1 > [2\sqrt{5}]^{-1}$.

Finite Horizon

The simplest case of Sinai billiard with finite horizon is obtained by employing two circular obstacles as in figure **??**. We have thus a torus of size one together with a circular obstacle at the point $(0, 0)$ with radius r_1 and a circular obstacle at the point $(1/2, 1/2)$ with radius r_2.[12] Clearly

$$r_1 + r_2 < \frac{1}{\sqrt{2}}$$

in order for the obstacels not to intersect each other. By the discussion of the infinite horizon case it follows that we can choose $r_1 > 1/(2\sqrt{2})$ and $0 < r_2 < 1/\sqrt{2} - r_1$ to have a Sinai billiard with disjoint obstacles and finite horizon. For example one could choose $r_1 = 3/7$ and $r_2 = 1/4$ as in figure **??**.

In the following we will need a more in depth undertanding of the above model. Let us consider a regularized Poincarè section of the type introduced in section 5.1.3 and discuss the structure of the singulatirity lines for such a section.

The first step is to understand multiple consecutive tangencencies. Let us strart with a double tangency of which the first is with the central obstacle. By simmetry one can limit the analysis to the case in which the second takes place with the upper right copy of the

[11] If this it is not clear, read again the discussion of polygonal billiards in section 5.0.1.

[12] Remember that the coordinates are in the universal covering of the torus and that the points $(1/2, -1/2)$, $(1/2, 1/2)$, $(-1/2, 1/2)$, $(-1/2, -1/2)$ are identified.

obstacle (see Figure ??. The position of the particle at time t is given by $r_1 n(\theta) + v(\theta)t$, where $n(\theta) = (\cos\theta, \sin\theta)$ and $v(\theta) = (\sin\theta, -\cos\theta)$, so we have the next two equations

$$\|r_1 n(\theta) + v(\theta)t - p\| = r_2$$
$$\langle r_1 n(\theta) + v(\theta)t - p, v(\theta)\rangle = 0,$$

where $p = (1/2, 1/2)$ are the coordiantes of the center of teh second obstacle (of course we are working in the universal covering). The first equation determine the value of t for which the second collision takes place while the second impose that the collision is tangent. Solving the above equations yields

$$\frac{1}{\sqrt{2}}\cos\left(\frac{\pi}{4} - \theta\right) = \langle n(\theta), p\rangle = r_1 \pm r_2.$$

Accordingly we have four solutions: $\frac{\pi}{4} - \theta = \pm(\cos^{-1}\sqrt{2}(r_1 \pm r_2))$. In fact, only two are really relevant since the other two are obtained by simmetry around the line joining the two centers. It remains to check that the above tajectories do not intersect any other obstacle between the two tangencies. In fact, it turns out that the trajectories of the type $\frac{\pi}{4} - \theta = \pm(\cos^{-1}\sqrt{2}(r_1 - r_2))$ have a tangent collision with the central scatterer before colliding with the corner one. It is then easy to see that there can be at most four consecutive tangencies after that the next collision will take place with an angle of more than 70 degree, see Figure ??.

5.5 Hard spheres

We have already mentioned several times that the motions of sevarl discks or balls that collide elastically amons themselves is an example of billiard. It is finally time to look at this interesting fact in detail. We will treat explicitely the case of two disck in two dimensions and we will comment on the more genral case.

5.5.1 Two balls, two dimensions

Let us start by considering the motion of two identical disks of mass one in $\mathbb{T}^2$. Again the motion is linear plus elastic collisions.

Clearly, the disks must have radius r sufficiently small in order to fit in the torus,for simplicity we assume $r < \frac{1}{4}$. The phase space is $X = \mathbb{T}^4 \times \mathbb{R}^4$.

If $x_1, x_2 \in \mathbb{T}^2$ are the coordinated of the center of the disks, the velocity changes at collision according to the law

$$\begin{cases} v_1^+ = v_1^- - \langle\eta, v_2^- - v_1^-\rangle(v_2^- - v_1^-) \\ v_2^+ = v_2^- + \langle\eta, v_2^- - v_1^-\rangle(v_2^- - v_1^-) \end{cases} \tag{5.5.11}$$

where η is a unit vector in the direction $x_2 - x_1$.[13]

Here there are more integral of motion than in the previous cases: the energy $E = \frac{1}{2}(\|v_1\|^2 + \|v_2\|^2)$ and the total momentum $P = v_1 + v_2$. Thus, if we want to obtain an ergodic

[13]To be precise $x_2 - x_1$ has no meaning since $\mathbb{T}^2$ it is not a linear space. Yet, at collision, the distance between the two disks is $2r$, so the global structure of $\mathbb{T}^2$ is irrelevant and we can safely confuse it with a piece of $\mathbb{R}^2$.

systems, we have to reduce the system vie the integral of motion. we will then consider that phase spaces

$$X_{E,P} = \{(x_1, x_2, v_1, v_2) \in X \mid \frac{1}{2}(\|v_1\|^2 + \|v_2\|^2) = E;\ v_1 + v_2 = P\}.$$

Since, in the velocity space, the previous conditions correspond to the intersection between the surface of a four dimensional sphere (S^3) and a two dimensional linear space, the velocity vectors $(v_1 + v_2)$ is contained in a one dimensional circle. Thus, topologically, $X_{E,P} = \mathbb{T}^4 \times S^1$.[14] It is then natural to choose an angle θ as coordinate on S^1, moreover, since

$$2E = \|v_1\|^2 + \|v_2\|^2 = \frac{1}{2}\|v_1 - v_2\|^2 + \frac{1}{2}\|P\|^2,$$

it is hard to resist setting $v_2 - v_1 = v(\theta)$.[15] Hence,

$$\begin{cases} v_1 = \frac{1}{2}(P - v(\theta)) \\ v_2 = \frac{1}{2}(P + v(\theta)). \end{cases}$$

The free motion is then given by

$$\begin{cases} x_1(t) = x_1(0) + \frac{1}{2}(P - v(\theta))t \\ x_2(t) = x_2(0) + \frac{1}{2}(P + v(\theta))t. \end{cases}$$

Accordingly,

$$\begin{cases} x_1(t) + x_2(t) = x_1(0) + x_2(0) + Pt \mod 1 \\ x_2(t) - x_1(t) = x_2(0) - x_1(0) + v(\theta)t \mod 1. \end{cases}$$

It is then clear the need to introduce the two new variables $Q = x_1 + x_2$ and $\xi = x_2 - x_1$. The variable Q performs a translation on the torus, such a motions is completely understood and we can then disregard it. The only relevant motion is the one in the variables (ξ, θ). The reduced phase space is then $\simeq \mathcal{B} \times S^1$ where $\mathcal{B} = \mathbb{T}^2 \backslash \{\|\xi\| \leq 2r\}$, that is the torus minus a disk of radius $2r$. The domain $\mathcal{B}$ is represented in Figure ?? and, apart the different choice of the fundamental domain, it corresponds exactly to the table of the simplest Sinai billiard.

The free motion corresponds to the free motion of a point as well, while at collision, from (5.5.11), we have

$$v(\theta^+) = v(\theta^-) - 2\langle \frac{\xi}{2r}, v(\theta^-)\rangle v(\theta^-)$$

that is exactly the elastic reflection from the disk! In conclusion the reduced system is exactly the simplest case of Sinai Billiard already studied in **??**.

Problems

5.1 Given a rectangular box $\mathcal{B}$, with its sides labeled by $\{1, 2, 3, 4\}$ and let $R_i(\mathcal{B})$ be the reflection with respect to the side i of the box $\mathcal{B}$.[16] Let R_0 be the identity. Consider $G = \cup_{n=0}^{\infty}\{0, 1, 2, 3, 4\}^n$, if $g \in G$ then we define $R_g(\mathcal{B}) = R_{g_1}(\cdots R_{g-n}(\mathcal{B})\cdots)$ and, for

[14] Of course, we are considering only the cases $E \neq 0$, $P \neq 0$.
[15] As usual $v(\theta) = (\sin\theta, \cos\theta)$.
[16] The label attached to the sides of the reflected boxed are the one obtained naturally from the old ones.

each $g^i \in \{0,1,2,3,4\}$, $i \in \{1,2\}$, $g = g^2 \circ g^1 \in \{0,1,2,3,4\}^{n_1+n_2}$, is defined by $g_k = g^1_k$ for $k \le n_1$ and $g_k = g^2_{n-k}$, for $k > n_1$. Verify that $R_{g^2}(R_{g^1}(\mathcal{B})) = R_g(\mathcal{B})$. Introduce the equivalence relation $g_1 \sim g_2$ iff $R_{g^1}(\mathcal{B}) = R_{g^2}(\mathcal{B})$. Let $\tilde{G}$ be the collection of the equivalence classes. Verify the $\tilde{G}$ is a commutative group with respect to the operation $\circ$. (hint: Note that the geometrical meaning is simply that the final position of the box after a certain number of reflections does not depend on the order of the reflections. Clearly it suffices to check such a property for two reflections.)

5.2 Study the motion in a triangular billiard when the angle defining the triangle are all rational multiple of π. (Hint: use reflections again)

5.3 Study the motion in an elliptical billiard. (Hint: Verify that there exist an integral of motion.)

5.4 Verify that the caustics correspond to a two dimensional torus. (Hint: ...)

5.5 Check that the maps ϕ^t generated by a billiard flow are symplectic. (Hint: It is obvious for the free flow, it remains to check it for the reflections. This can be done by using formulae ???)

5.6 Find a change of variable that transforms the symplectic form in a regularized boundary section in the standard symplectic form.

5.7 Verify that, in a regularized boundary section, the map is continuous accros a syngularity line corresponding to a tangency.

5.8 Prove that, given an $n \times n$ matrix A the norm $\|A\| := \sup_{v\in\mathbb{R}^n} \frac{\|Av\|}{\|v\|}$ where $\|v\| := \sqrt{\sum_n v_n^2}$ satisfies

$$\|A\| \le \text{constant} \max_{ij} |A_i j|$$

and compute explicitely the optimal constant.

5.9 Determine the rate at wich the second derivative explode as one gets near to a tangency singularity in the Sinai billiard with one circular obstacle in the torus.

5.10 Compute the number of collisions in a convex angle.

5.11 Show that in a Sinai billiard on $\mathbb{T}^2$ for which there exists trajectories that spend an arbitrary long time without colliding there must exists trajectories that never collide. (Hint: some continuity...)

5.12 Let

$$L_i = \begin{pmatrix} 1 & 0 \\ t_i & 1 \end{pmatrix} \quad \text{and} \quad R_i = \begin{pmatrix} 1 & k_i \\ 0 & 1 \end{pmatrix},$$

for each $u \in \mathbb{R}^+$ write

$$\prod_{i=0}^{n} R_i L_i \begin{pmatrix} 1 \\ u \end{pmatrix} = \lambda_n \begin{pmatrix} 1 \\ u_n \end{pmatrix}.$$

Show that

$$u_n = k_n + \cfrac{1}{t_n + \cfrac{1}{k_{n-1} + \cfrac{1}{t_{n-1} + \ddots + \cfrac{1}{t_1 + \frac{1}{k_1 + \frac{1}{u}}}}}}$$

And find conditions for the convergence of the continuous fraction. (Hint: see Problem 4.29)

5.13 Study two disk with different masses.

5.14 Prove that the Poincarè map for the Sinai billiard is piecewise Hölder of Hölder exponent $\frac{1}{2}$.

Bibliography

[1] J.Aaronson *An introduction to infinite ergodic theory* American Mathematical Society, Mathematical Surveys and Monographs, **50** (1997)

[2] D.V. Anosov *Geodetic flows on closed Riemannian manifolds with negative curvature* Proc.Inst.Steklov **90**, 1–235 (1967)

[3] V.I.Arnold*Mathematical Methods in Celestial Mechanics* Springer Verlag, Berlin, Heidelgerg, New York, (1978, 1989)

[4] R.Artuso, G.Casati, I.Guarneri *Geometric scaling of correlation decay in chaotic billiards* Phys. Rev. E (3) **51**, no. 5, part A, R3807–R3810, (1995).

[5] R.Artuso, G.Casati, I.Guarneri *Numerical experiments on billiards* J. Statist. Phys. **83**, no. 1-2, 145–166, (1996).

[6] R.Artuso, G.Casati, I.Guarneri *Numerical study on ergodic properties of triangular billiards* Phys. Rev. E (3) **55**, no. 6, part A, 6384–6390, (1997).

[7] V.I.Arnold, A.Avez *Problèmes ergodiques de la mécanique clasique* Gauthier-Villards Paris (1967)

[8] D.V. Anosov, Ya.G.Sinai *Certain smooth ergodic systems* Russ. Math. Surv. **22**, 103 – 167 (1982)

[9] Garrett Birkhoff. *Extensions of Jentzsch's theorem.* Trans. Amer. Math. Soc., 85:219–227 (1957).

[10] Garrett Birkhoff. *Lattice theory*, volume **25** of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, R.I., third edition (1979).

[11] Boothby *Introduction to differential geometry*

[12] L.A. Bunimovich *On billiards close to dispersing* Math. USSR *sl* 95, 295–312 (1979)

[13] L. A. Bunimovich *On the ergodic properties of nowhere dispersing billiards* Comm.-Math.Phys. **65** 295 – 312 (1979)

[14] L.A. Bunimovich *Dynamical Systems of Hyperbolic Type with Singularities*–in *Dynamical Systems II* Ya.G.Sinai ed., Encyclopaedia of Mathematical Sciences, Springer–Verlag, Berling, Heidelberg, New York, 151–178 (1989)

[15] L.A. Bunimovich, C. Liverani, S. Pellegrinotti, Y.Suhov *Ergodic Systems of n Balls in a Billiard Table* Communications in Mathematical Physics**146** 357–396 (1992)

[16] L.A. Bunimovich, J.Rehacek, *Nowhere dispersing* 3*D billiards with non-vanishing Lyapunov exponents* Communications in Mathematical Physics **189**, no. 3, 729–757, (1997).

[17] L.A. Bunimovich, J.Rehacek, *On the ergodicity of many-dimensional focusing billiards. Classical and quantum chaos* Ann. Inst. H. Poincaré Physique Théorique **68**, no. 4, 421–448, (1998).

[18] L.A.Bunimovich, Ya.G.Sinai, *Markov partition for dispersed billiards* Comm. Math. Phys. **78**, no. 2, 247–280, (1980/81)

[19] L.A.Bunimovich, Ya.G.Sinai, *Erratum: "Markov partition for dispersed billiards"* [Comm. Math. Phys. **78** (1980/81), no. 2, 247–280; MR 82e:58059]. Comm. Math. Phys. **107**, no. 2, 357–358, (1986).

[20] L.A.Bunimovich, Ya.G.Sinai, N.I. Chernov *Markov partitions for two-dimensional hyperbolic billiards* (Russian) Uspekhi Mat. Nauk **45** (1990), no. 3(273), 97–134, 221; translation in Russian Math. Surveys **45**, no. 3, 105–152, (1990).

[21] L.A.Bunimovich, Ya.G.Sinai, N.I. Chernov *Statistical properties of two-dimensional hyperbolic billiards* (Russian) Uspekhi Mat. Nauk **46** (1991), no. 4(280), 43–92, 192; translation in Russian Math. Surveys **46**, no. 4, 47–106, (1991).

[22] D.Burago, S. Ferleger, A.Kononenko *Uniform estimates on the number of collisions in semi-dispersing billiards* Annals of Mathemathics (2) **147**, no. 3, 695–708, (1998).

[23] D.Burago, S. Ferleger, A.Kononenko *Unfoldings and global bounds on the number of collisions for generalized semi-dispersing billiards* Asian J. Math. **2** , no. 1, 141–152, (1998).

[24] K. Burns, M. Gerber *Continuous invariant cone families and ergodicity of flows in dimension three* Erg.Th.Dyn.Syst. **9** 19–25, (1989)

[25] H.Busemann. P.J.Kelly *Projective Geometry and Projective metrics* Academic Press, New York (1953)

[26] N.I.Chernov, Ya.G.Sinai *Ergodic properties of some systems of* 2*-dimensional discs and* 3*-dimen- sional spheres* Russ.Math.Surv. **42**, 181–207, (1987)

[27] Mark Demers, C.Liverani, Niloofar Kiamari. *Transfer operators in Hyperbolic Dynamics An introduction.* 33 Colloquio Brasilero de Matematica. Editora do IMPA. pp.252 (2021).

[28] V.J.Donnay *Using integrability to produce chaos: billiards with positive entropy* Comm. Math.Phys. **141**, 225–257 (1991)

[29] V.J.Donnay, *Elliptic islands in generalized Sinai billiards* Ergodic Theory Dynam. Systems **16**, no. 5, 975–1010, (1996).

[30] H.R.Dullin *Linear stability in billiards with potential* Nonlinearity **11**, no. 1, 151–173, (1998).

[31] do Carmo, Manfredo Perdigão, *Riemannian geometry.* Translated from the second Portuguese edition by Francis Flaherty. Mathematics: Theory and Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.

[32] G. Gallavotti *Modern theory of billiards—an introduction. Global analysis and its applications* (Lectures, Internat. Sem. Course, Internat. Centre Theoret. Phys., Trieste, 1972), Vol. II, pp. 193–202. Internat. Atomic Energy Agency, Vienna, (1974).

[33] G. Gallavotti *Lectures on the billiard* in Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle., Wash., 1974), pp. 236–295. Lecture Notes in Phys., Vol. **38**, Springer, Berlin, (1975).

[34] G. Gallavotti *The Elements of Mechanics*, Text and Monograph in Physics, Springer-Verlag, New-York, Heidelberg, Berlin, (1983).

[35] G. Gallavotti *Aspetti della teoria ergodica e statistica del moto* Quaderni dell'Unione Matematica Italiana, **21** Pitagora Editrice, Bologna (1981)

[36] G.Gallavotti, D.S.Ornstein *Billiards and Bernoulli schemes* Comm. Math. Phys. **38**, 83–101, (1974).

[37] P.Garrido *Kolmogorov-Sinai entropy, Lyapunov exponents, and mean free time in billiard systems* J. Statist. Phys. **88**, no. 3-4, 807–824, (1997).

[38] P. Garrido, G.Gallavotti *Billiards correlation functions* J. Statist. Phys. **76**, no. 1-2, 549–585, (1994).

[39] Hirsch, M. W.; Pugh, C. C.; Shub, M. *Invariant manifolds.* Lecture Notes in Mathematics, Vol. **583**. Springer-Verlag, Berlin-New York, (1977).

[40] E. Hopf *Statistik der Geodätischen Linien in Mannigfaltigkeiten Negativer Krümmung* Ber. Verh. Sächs. Akad. Weiss., Leipzig **91**, 261–304, (1939)

[41] E. Hopf *Statistik der Lösungen Geodätischen Probleme vom unstabilen Typus. II* Mathematische Annalen **117**, 590–608 (1940)

[42] A. Katok in collaboration with K. Burns *Infinitesimal Lyapunov functions, invariant cone families and stochastic properties of smooth dynamical systems* preprint (1992)

[43] A.Katok, B.Hasselblatt *Introduction to the Modern Theory of Dynamical Systems* Camdridge University Press, Cambridge, (1995)

[44] A.Knauf *On soft billiard systems* Phys. D **36**, no. 3, 259–262, (1989).

[45] A. Katok, J.-M. Strelcyn with the collaboration of F. Ledrappier and F. Przytycki *Invariant manifolds, entropy and billiards; smooth maps with singularities* Lecture Notes in Math. **1222** Springer-Verlag (1986)

[46] A. Krámli, N. Simányi, D. Szász *A "Transversal" Fundamental Theorem for Semi-Dispersing Billiards* Communications in Mathematical Physics **129** 535–560 (1990), (see also Erratum)

[47] A. Krámli, N. Simányi, D. Szász *Erratum: "A 'Transversal' fundamental theorem for semi-dispersing billiards"* [Comm. Math. Phys. **129**, no. 3, 535–560, (1990)]. Comm. Math. Phys. **138**, no. 1, 207–208, (1991).

[48] A. Krámli, N. Simányi, D. Szász *K–property of three Billiard Balls* Annals of Mathematics **133**, 37–72,(1991)

[49] A. Krámli, N. Simányi, D. Szász *K–property of four Billiard Balls* Comm.Math.Phys. **144** 107–148, (1992)

[50] F.Ledrappier,*Propriètè ergodiques des mesures de Sinai*, C.R.Acad.Sci., Paris, Sèr. I, **294**, 593–595 (1982).

[51] Y.-E.Lévy *A note on Sinai and Bunimovich's Markov partition for billiards* J. Statist. Phys. **45**, no. 1-2, 63–68, (1986).

[52] E.H.Lieb, M.Loss *Analysis* Graduate studies in Mathematics, vol. **14**, American Mathematical Society (1996)

[53] C.Liverani, M.Wojtkowski *Generalization of the Hilbert metric to the space of positive definite matrices* Pacific Journal of Mathematics **166**, n. 2, 339–355 (1994)

[54] C.Liverani, M.Wojtkowski *Ergodicity in Hamiltonian systems* Dynamics Reported **4** (1995)

[55] R. Mañe *Ergodic Theory an differentiable dynamics* Spring-Verlag (1987)

[56] Roger D. Nussbaum. *Hilbert's projective metric and iterated nonlinear maps.* Mem. Amer. Math. Soc., **75** (391):iv+137, (1988).

[57] V. I. Oseledets *A Multiplicative Ergodic Theorem: Characteristic Lyapunov Exponents of Dynamical Systems* Trans. Moscow Math. Soc. **19**, 197–231, (1968)

[58] Ya. B. Pesin *Lyapunov Characteristic Exponents and Smooth Ergodic Theory* Russ. Math. Surveys **32**, 4, 55–114 (1977)

[59] H.L.Royden *Real Analysis*, III edition, Macmillan Publishng Company, New York, Collier Macmillan Publishing, London (1988)

[60] M.Reed, B.Simon *Funtional Analysis, I*, Revised and enlarged edition, Accademic Press, New York,London,Toronto,Sydney,San Francisco (1980)

[61] W. Rudin *Real and Complex Analysis*, Mc Graw-Hill series in higher Mathematics, Second Edition (1974).

[62] Ya.G.Sinai *Markov partitions and C-diffeomorphisms* Funct. Anal. Appl. **2**, 61–82, (1968)

[63] Ya.G.Sinai *Dynamical systems with elastic reflections* Russ.Math.Surveys **25** 137 – 189 (1970)

[64] Ya.G. Sinai (Ed.)*Dynamical system II* Springer- Verlag (1989)