Alma Mater Studiorum · University of Bologna

**School of Science**
**Department of Physics and Astronomy**
**Master Degree in Physics**

# On the Kolmogorov property of a class of infinite measure hyperbolic dynamical systems

**Supervisor:**
**Prof. Marco Lenci**

**Submitted by:**
**Giovanni Canestrari**

Academic Year 2020/2021

*Sono tornato là*
*dove non ero mai stato.*
*Nulla, da come non fu, è mutato.*
*Sul tavolo (sull'incerato*
*a quadretti) ammezzato*
*ho ritrovato il bicchiere*
*mai riempito. Tutto*
*è ancora rimasto quale*
*mai l'avevo lasciato.*

Giorgio Caproni, Ritorno

# Sommario

Le mappe lisce con singolarità descrivono importanti fenomeni fisici come l'urto tra sfere rigide tra loro e/o con le pareti di un contenitore. Le domande sulle proprietà ergodiche di questi sistemi (che possono essere mappati in sistemi di tipo biliardo) furono sollevate per primo da Boltzmann nel diciannovesimo secolo e pongono i fondamenti della Meccanica Statistica. I sistemi di tipo biliardo descrivono anche i moti diffusivi di elettroni che urtano contro nuclei di carica positiva (gas di Lorentz) e in questo caso la misura fisica preservata dalla dinamica può essere considerata a tutti gli effetti come infinita. E' dunque di grande importanza studiare le proprietà ergodiche di mappe che preservano una misura infinita. Lo scopo di questa tesi è di presentare un risultato originale sulle mappe liscie con singolarità che preservano una misura infinita. Tale risultato dimostra l'atomicità della $\sigma$-algebra di coda (e dunque forti proprietà stocastiche) in presenza di un comportamento totalmente conservativo.

# Abstract

Smooth maps with singularities describe important physical phenomena such as the collisions of rigid spheres among them and/or with the walls of a container. Questions about the ergodic properties of these models (which can be mapped into billiard models) were first raised by Boltzmann in the nineteenth century and lie at the foundation of Statistical Mechanics. Billiard models also describe the diffusive motion of electrons bouncing off positive nuclei (Lorentz gas models) and in this situation the physical measure can be considered infinite. It is therefore of great importance to study the ergodic properties of maps when the measure they preserves is infinite. The aim of this thesis is to present an original result on smooth maps with singularities which preserve an infinite measure. Such result establishes the atomicity of the tail $\sigma$-algebra (and hence strong chaotic properties) in the presence of a totally conservative behavior.

# Contents

# Chapter 1

# Introduction

In this thesis we present an original result that states that certain billiard-like systems (smooth, hyperbolic maps with singularities) enjoy strong chaotic properties, even when they have *infinite measure*. In particular, we prove that these dynamical systems have an *atomic tail $\sigma$-algebra* with respect to their partition into stable manifolds. This result has been proved for finite measure billiard-like systems by Pesin [Pes1,Pes2], but his methods seem unable to be easily carried over to the infinite measure case because of the use of entropy theory therein. Billiards are mathematical models for many physical systems where one or more particles collide with the walls of a container and/or with each other. The dynamical properties of such models are determined by the shape of the walls of the container, and may vary from completely regular (integrable) to fully chaotic. The most intriguing, though least elementary, are chaotic billiards. They include the classical models of *hard balls* studied by L. Boltzmann in the nineteenth century, the Lorentz gas introduced to describe electricity in 1905, as well as modern dispersing billiard tables due to Ya. Sinai. The mathematical theory of chaotic billiards was born in 1970 when Ya. Sinai published his seminal paper [S], and is now 50 years old. But during these years it has grown and developed at a remarkable speed and has become a well established and flourishing area within the modern theory of dynamical systems and statistical mechanics. The aim of this introduction is to explain why a physicist or a mathematician may be interested in the dynamics of billiards. We have already mentioned two applications, namely the Lorentz gas [Lo] and the hard ball systems. To explain a little better what these systems have in common with billiards we shortly define billiard dynamics. We anticipate here some definitions that will be studied in greater detail later.

**Definition 1.0.1.** Let $\mathcal{D} \subset \mathbb{R}^2$ be a domain with smooth or piece-wise smooth boundary. A billiard system corresponds to a free motion of a point particle inside $\mathcal{D}$ with specular reflections off the boundary $\partial\mathcal{D}$.

The definition of billiard dynamics mirrors almost perfectly the physics of the *billiard*

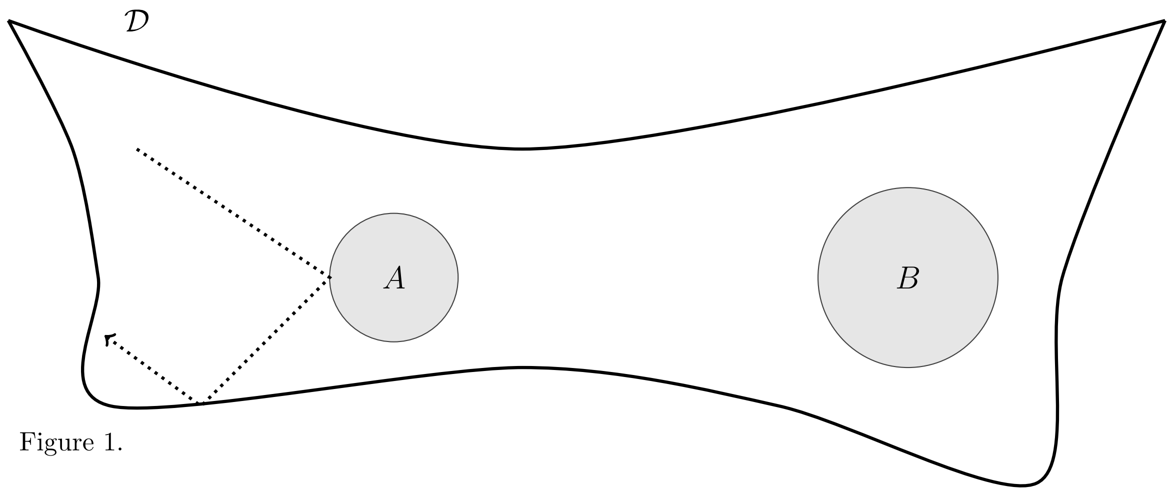


Figure 1.

*game* the reader is certainly familiar with. The main differences being two: the billiard ball is point-like and moves without friction for an infinite time (except when it hits a corner and ends up in the hole) *and* the billiard table is not necessarily rectangular-shaped but can assume a great variety of shapes. A billiard table may look, for example, like this:

The thick black line, which defines the domain $\mathcal{D} \subset \mathbb{R}^2$, represent the wall of the container. When the trajectory of the ball (dotted line) meets them, it bounces off following the classical rule 'angle of incidence is equal to the angle of reflection'. The two gray disks $A$ and $B$ are called *scatterers*. Usually, they are not included in $\mathcal{D}$ by definition, but behave exactly as the wall: the ball bounces off them with the usual rule. When they are present the system resembles more a *flipper* than a billiard. We will see later that the configuration space of two hard-balls colliding on a torus maps exactly into a billiard system with a point-like particle bouncing elastically off the walls.

## 1.1 Lorentz gas

One can also consider $\mathcal{D} \subset \mathbb{T}^2$ instead as $\mathcal{D} \subset \mathbb{R}^2$. In this case $\mathcal{D} = \mathbb{T}^2 \setminus \cup \mathbb{O}_i$, where $\mathbb{O}_i$ are the scatterers. Convex billiards defined on the torus are called Sinai billiards[1], in honor of Ya. Sinai who first carried a detailed study of them [S]. Upon lifting[2] the system to $\mathbb{R}^2$, the motion of the billiard ball in this infinite flipper exactly defines the dynamics of a periodic Lorentz gas (PLG), in which an electron bounces off the positive

[1] For a more precise definition of Sinai billiard see the next sections.

[2] Leaving aside the precise definition of 'lifting', it simply means that we "tessellate" the plane $\mathbb{R}^2$ with infinite identical copies of our fundamental domain $\mathbb{T}^2$ (and of course along with it we copy everything that stays inside the fundamental domain).

nuclei.

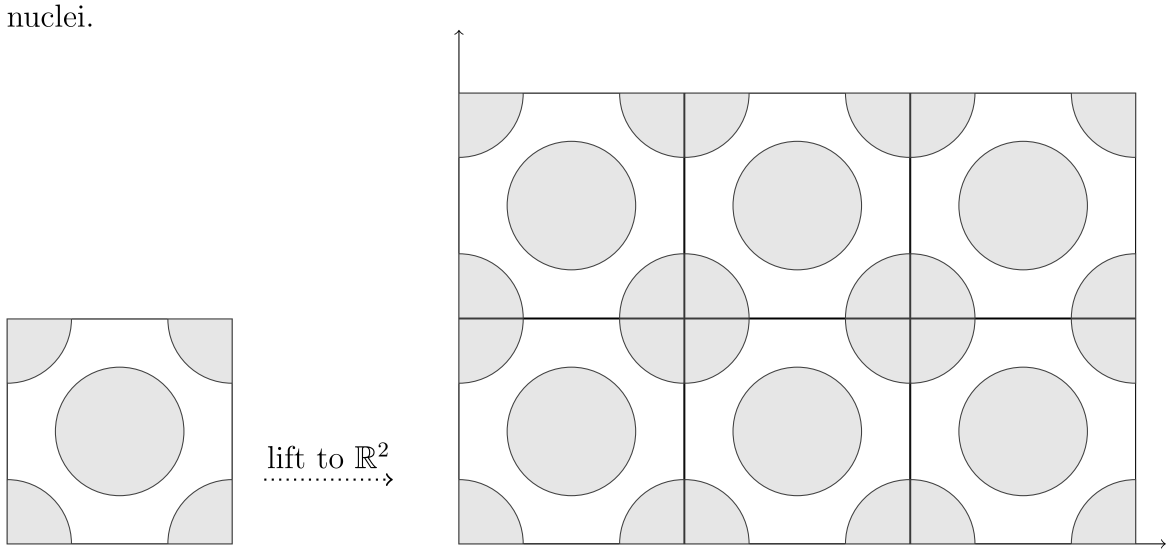


Figure 2.

Note that the Lorentz gas (which is the lifted version in $\mathbb{R}^2$) is similar to the billiard on the torus. A remarkable difference between the two models is that the 'unlifted' version has finite measure while the lifted one infinite. Nonetheless, one tries to infer statistical properties of the Lorentz gas studying the 'smaller' system on the torus, which is a canonical example of chaotic billiard. Indeed, the original part of this thesis goes in this direction. In this program, a key property that has to be satisfied by the lifted system with infinite measure is that of conservativity almost everywhere. Roughly speaking, every electron that starts its motion from one of the scatterer has to retrace itself and pass close to where it started infinitely many times in the future. This property is also called *recurrence* and expresses the concept that history repeats itself. For periodic Lorentz gas the problem of recurrence has baffled mathematicians for decades, and was answer in the affirmative only in the last few years, by Schmidt [Sch] and Conze [Co] independently. A further progress has been achieved by Lenci [L3, L4] who proved the recurrence for a certain class of aperiodic Lorentz gases. Therefore, it is of great interest to better understand the implications of recurrence, in terms of the statistical properties of the dynamics when it comes along with others properties of the map.

## 1.2 Hard balls

Here we describe a mechanical system of two hard disks that reduces to a dispersing billiard. This mechanical model is actually a prototype of all dispersing billiard tables and is the model that motivated Sinai to introduce dispersing billiards after he studied the motion of hard disks and hard balls in the 1960s.

Consider two identical disks of unit mass $m = 1$ and a small radius $r$ moving on the unit

torus $\mathbb{T}^2$ without obstacles or walls of any kind. The disks move freely with constant velocities and collide with each other elastically, as described below. Let the disks collide and denote their velocity vectors before the collisions by $u_-$ and $v_-$. Let $L$ denote the common tangent line to the disks at the moment of collision. For every vector $v$ denote by $v^\perp$ and $v^{||}$ its components perpendicular and parallel to the line $L$, respectively. We call $v^{||}$ the *tangential* component of $v$ and $v^\perp$ the *normal* component of $v$.
According to the law of elastic collision, the postcollisional velocity vectors of the disks are

$$u_+ = u_-^{||} + v_-^\perp \quad \textit{and} \quad v_+ = v_-^{||} + u_-^\perp. \tag{1.2.1}$$

In other words, *the velocity vectors keep their tangential components but exchange their normal components*. Observe that this rule implies the preservation of the total kinetic energy

$$||u_+||^2 + ||v_+||^2 = ||u_-||^2 + ||v_-||^2$$

and the total momentum

$$u_+ + v_+ = u_- + v_-$$

(recall that the disks have unit mass). The conservation of the total momentum implies that the center of mass $(x_c, y_c)$ moves on the torus with a constant velocity vector. This is a simple periodic or quasi periodic motion, and we eliminate it from the picture assuming that the total momentum is zero; i.e. $u + v = 0$ at all the times.
Now the center of mass is at rest. More precisely, let $0 \le x, y \le 1$ denote rectangular coordinates on $\mathbb{T}^2$ and $(x_1, y_1)$ and $(x_2, y_2)$ the centers of our disks. Choose the coordinate systems so that, initially, the center of mass

$$x_c = \frac{1}{2}(x_1 + x_2) \quad \textit{and} \quad y_c = \frac{1}{2}(y_1 + y_2)$$

coincides with the center of the torus $(\frac{1}{2}, \frac{1}{2})$; equivalently,

$$x_1 + x_2 = 1 \quad \textit{and} \quad y_1 + y_2 = 1. \tag{1.2.2}$$

Clearly, since the total momentum is zero, the position of the center of mass remains constant and Equation (1.2.2) will hold at all times. The coordinates $x$ and $y$ are cyclic, so they can occasionally increase or decrease by one; therefore, when $x_1$ increases by one, for example, then $x_2$ must decrease by one. Next, the disks collide with each other whenever the distance between their centers is $2r$, i.e. whenever

$$(x_1 - x_2 + p)^2 + (y_1 - y_2 + q)^2 = (2r)^2$$

for some $p = 0, \pm 1$ and $q = 0, \pm 1$. Due to (1.2.2), this is equivalent to

$$\left(x_1 - \frac{1-p}{2}\right)^2 + \left(y_1 - \frac{1-q}{2}\right)^2 = r^2. \tag{1.2.3}$$

Simple substitution of the possible values of $p, q \in \{0, \pm 1\}$ shows that Equation (1.2.3) specifies four circles of radius $r$ with centers at $(0,0), (0, \frac{1}{2}), (\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$. Let $\mathbb{O}_i$ denotes the circles (1.2.3). We now examine the transformation of the velocity vectors $u$ and $v$ at collisions. Since $u + v = 0$, we obviously have $u_-^{||} = -v_-^{||}$ and $u_-^{\perp} = -v_-^{\perp}$ in the notation of (1.2.1); therefore

$$u_+ = u_-^{||} - u_-^{\perp} \quad \textit{and} \quad v_+ = v_-^{||} - v_-^{\perp}.$$

In other words, each velocity vector simply gets reflected off the common tangent line $L$ of the colliding disks. Observe that $L$ is parallel to the tangent line to the circle $\mathbb{O}_i$ on which the point $(x_1, y_1)$ lies at the moment of collision. Also note that $||u|| = ||v|| = \frac{\sqrt{2}}{2}$ remain constant in time.

Therefore, the center $(x_1, y_1)$ of the first disk moves on $\mathbb{T}^2$ with a constant velocity vector until it hits one of the circles $\mathbb{O}_i$. Then its velocity gets reflected across the tangent line to that circle at the point of hit, and so on. This is exactly a billiard trajectory[3] on the table $\mathbb{T}^2 \setminus \cup \mathbb{O}_i$.

[3]To be precise, the point $(x_1, y_1)$ moves at a constant, but not unit, speed. One can rescale the time to ge a conventional unit speed.

# Chapter 2

# Some notions of infinite ergodic theory

In this chapter we deal with the property of *conservativity*. While for *finite measure measure preserving transformations* conservativity is just a corollary (it is the content of a theorem by Poincaré), in the case of infinite measure conservativity is not assured and it gains a central role in the development of the theory. Here, the aim is just to define this property and characterize it a little bit but we will always stay quite *near the surface* of the infinite ergodic theory. A great part of this summary owes to the first chapter of the book of Aaronson [A], which is a very detailed introduction to the subject. Let us start by describing the very basics of ergodic theory and by stating the Poincaré theorem for the finite measure case.
A measure space is the triple $(X, \mathfrak{A}, \mu)$ where $X$ is a set, $\mathfrak{A}$ is a $\sigma$-algebra on $X$ and $\mu$ is a measure on $(X, \mathfrak{A})$. A measure space $(X, \mathfrak{A}, \mu)$ is said *probability space* if $\mu(X) = 1$. Often, it is required that the measure space is *nice enough* by asking it to be a *standard measure space.*

**Definition 2.0.1.** A *Polish space* is a complete, separable metric space. Let $X$ be a Polish space. The collection of *Borel sets* $\mathfrak{B}(X)$ is the $\sigma$-algebra of subsets of $X$ generated by the collection of open sets.

**Definition 2.0.2.** A *standard measurable space* (or standard Borel space) $(X, \mathfrak{B})$ is a Polish space $X$ equipped with its Borel sets.

**Definition 2.0.3.** A *standard measure space* is a measure space $(X, \mathfrak{B}, m)$ where $(X, \mathfrak{B})$ is a standard measurable space.

Standard measure spaces can be regarded as the bridge from topological notions and measure-theoretic ones, in that they assign positive measure to the open sets.

**Definition 2.0.4.** Let $(X, \mathfrak{B}, m)$ and $(X', \mathfrak{B}', m')$ be measure spaces, and let $A \in \mathfrak{B}$, $A' \in \mathfrak{B}'$. The map $T : A \to A'$ is *measurable* if $T^{-1}C \in \mathfrak{B}\ \forall C \in \mathfrak{B}'$. The measurable map $T : A \to A'$ is called *invertible* on $B \in \mathfrak{B} \cap A$ if $T$ is 1-1 on $B$, $TB \in \mathfrak{B}'$, and $T^{-1} : TB \to B$ is measurable.

Standard ergodic theory studies *measure preserving maps* in probability spaces. An extension of the notion of preservation of measure is *non-singularity.*

**Definition 2.0.5.** The measurable map $T : A \to A'$ is called (two-sided) *non-singular* if for $C \in \mathfrak{B}' \cap A'$, one has

$$m(T^{-1}C) = 0 \Longleftrightarrow m(C) = 0.$$

and *measure preserving* if $m(T^{-1}C) = m(C)$ for $C \in \mathfrak{B}' \cap A'$.

## 2.1 Recurrence and conservativity

One property that is enjoyed by all measure-preserving transformations in probability spaces is recurrence:

**Theorem 2.1.1.** *(Poincaré's Recurrence Theorem) Let $T : X \to X$ be a measure preserving transformation of a probability space $(X, \mathfrak{B}, m)$. Let $E \in \mathfrak{B}$ with $m(E) > 0$. Then almost all points of $E$ return infinitely often to $E$ under positive iterations by $T$ (i.e., there exists $F \subseteq E$ with $m(F) = m(E)$ such that for each $x \in F$ there is an infinite sequence $n_1 < n_2 < n_3 < ...$ of natural numbers with $T^{n_i}(x) \in F$ for each $i$).*

*Proof.* For $N \geq 0$ let $E_N = \bigcup_{n=N}^{\infty} T^{-n}E$. Then $\bigcap_{N=0}^{\infty} E_N$ is the set of all points of $X$ which enter $E$ infinitely often under positive iterations of $T$. Hence the set $F = E \cap \bigcap_{N=0}^{\infty} E_N$ consists of all points of $E$ that enter $E$ infinitely often under positive iterations by $T$. If $x \in F$ then there is a sequence $n_1 < n_2 < n_3 < ...$ of natural numbers with $T^{n_i}(x) \in E$ for all $i$. For each $i$ we have also $T^{n_i}(x) \in F$ because $T^{n_j - n_i}(T^{n_i}x) \in E$ for all $j$. It remains to show $m(F) = m(E)$.
Since $T^{-1}E_N = E_{N+1}$ we have $m(E_N) = m(E_{N+1})$ and hence $m(E_0) = m(E_N)$ for all $N$. Since $E_0 \supseteq E_1 \supseteq E_2 \supseteq ...$ we have $m(\bigcap_{N=0}^{\infty} E_N) = m(E_0)$. Therefore $m(F) = m(E \cap E_0) = m(E)$ since $E \subseteq E_0$. □

**Remark 2.1.2.**

1. In the proof the measure-preserving property of $T$ was used only in the weaker form of incompressibility (i.e., if $B \in \mathfrak{B}$ and $T^{-1}B \subseteq B$ then $m(B) = m(T^{-1}B)$). Therefore Theorem 2.1.1 is true for incompressible transformations.

2. Theorem 2.1.1 is false if a measure space of infinite measure is used. An example is given by the map $T(x) = x + 1$ defined on $\mathbb{Z}$ with the measure on $\mathbb{Z}$ which gives each integer unit measure. For this example if $E$ denotes the set $\{0\}$ then the set $E_N$ all have infinite measure but $\bigcap_{N=0}^{\infty} E_N$ is empty.

That property, that holds automatically for measure-preserving transformations on probability spaces, is called *recurrence*.

**Definition 2.1.3.** We say that a measure-preserving map $T$ on a (possibly infinite) measure space $(X, \mathfrak{B}, m)$ is *recurrent* if for every positive measure set $A \in \mathfrak{B}$ almost every $x \in A$ has the property that $T^{n_i}x \in A$ for an infinite sequence of natural numbers $n_1 < n_2 < n_3 < \dots$ .

An equivalent definition is the one in [A]:

**Definition 2.1.4.** A non singular transformation $T$ on $(X, \mathfrak{B}, m)$ is *recurrent* if

$$\liminf_{n\to\infty} |h \circ T^n - h| = 0 \quad a.e. \ \forall h : X \to \mathbb{R} \ \text{ measurable.}$$

These two definitions are equivalent. Indeed, the second implies the first simply by choosing $h = 1_A$ for any $A \in \mathfrak{B}$ of positive measure (we write $A \in \mathfrak{B}_+$); the fact that the second definition implies the first will be clear after the prosecution of our discussion. We begin by studying the extremely non-recurrent behavior exhibited by *wondering sets*. Let $T$ be a non-singular transformation of the standard measure space $(X, \mathfrak{B}, m)$.

**Definition 2.1.5.** A set $W \subset X$ is called a *wondering set* (for $T$) if the sets $\{T^{-n}W\}_{n=0}^{\infty}$ are disjoint. Let $\mathcal{W} = \mathcal{W}(T)$ denote the collection of measurable wondering sets. Note that $T^{-1}\mathcal{W}(T) \subseteq \mathcal{W}(T)$.

The next result dichotomize, and localize the recurrent behavior of a non-singular transformation. Denote by $\mathfrak{B}_+$ the sets in $\mathfrak{B}$ with positive measure.

**Theorem 2.1.6.** *(Halmos' recurrence theorem) Suppose that $A \in \mathfrak{B}$, $m(A) > 0$, then*

$$m(A \cap W) = 0 \ \forall W \in \mathcal{W} \Longleftrightarrow \sum_{n=1}^{\infty} 1_B \circ T^n = \infty \ \text{ on } B \ \forall B \in \mathfrak{B}_+ \cap A.$$

*Proof.* $(\Rightarrow)$.
Assume that $B \in \mathfrak{B}$ and $m(B \cap W) = 0 \ \forall W \in \mathcal{W}$. Let

$$B^- = \bigcup_{n=1}^{\infty} T^{-n}B.$$

Clearly $T^{-1}B^- \subseteq B^-$. We claim that $W_B := B \setminus B^-$ is a wondering set for $T$. To see this, we note that $W_B \subseteq T^{-n}B^c \subseteq T^{-n}W_B^c \ \forall n \geq 1$. whence $T^{-n}W_B \cap T^{-m}W_B = \emptyset$ $\forall n \neq m$. Since $W_B \in \mathcal{W}$, $W_B \subset B$, we have (by assumption) that $m(W_B) = 0$. Thus

$$B \subseteq B^- \qquad \mod m.$$

By the non-singularity of $T$, $\forall n \geq 1$, $m(T^{-n}W_B) = 0$ and so

$$T^{-n}B \subseteq T^{-n}B^- \qquad \text{mod } m,$$

whence ( mod $m$),

$$B \subseteq B^- = T^{-1}B^- = T^{-2}B^- = \ldots = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} T^{-n}B = \{\sum_{n=1}^{\infty} 1_B \circ T^n = \infty\}.$$

($\Leftarrow$).
Conversely, if $\exists W \in \mathcal{W}$ so that $m(W \cap A) > 0$, then clearly

$$\sum_{n=1}^{\infty} 1_B \circ T^n = 0 \text{ on } B := W \cap A \in \mathfrak{B}_+.$$

□

Now we would like to define maximal sets in which there is wondering and recurrent behavior. To do so we have to face the problem that a union of arbitrary many measurable sets (such as the wondering sets) is not necessarily measurable. It turns out that in the case of wondering sets this is not a real problem because $\mathcal{W}(T)$ is what is called an *hereditary collection.*

**Definition 2.1.7.** Let $(X, \mathfrak{B}, m)$ be a measure space. A collection $\mathfrak{H} \subseteq \mathfrak{B}$ is called *hereditary* if

$$C \in \mathfrak{H}, B \subset C, B \in \mathfrak{B} \Longrightarrow B \in \mathfrak{H}.$$

The hereditary collections have *measurable unions.*

**Definition 2.1.8.** A set $U \in \mathfrak{B}$ is said to *cover* the hereditary collection $\mathfrak{H}$ if $A \subseteq U$ mod $m$ $\forall A \in \mathfrak{H}$.
A hereditary collection $\mathfrak{H} \subseteq \mathfrak{B}$ is said to *saturate* $U \in \mathfrak{B}$ if $\forall B \in \mathfrak{B}$, $B \subseteq U$, $m(B) > 0$, $\exists C \in \mathfrak{H}$, $m(C) > 0$, $C \subseteq B$.
The set $U \in \mathfrak{B}$ is called *measurable union* of the hereditary collection $\mathfrak{H} \subseteq \mathfrak{B}$ if it both covers and is saturated by $\mathfrak{H}$.

A good property of a measurable union is that there is no more than one for each hereditary collection. To see this, let $U$ and $U' \in \mathfrak{B}$ be measurable unions of the hereditary collection $\mathfrak{H}$ and suppose that $m(U \setminus U') > 0$. Then, (since $\mathfrak{H}$ saturates $U$) there exists $C \in \mathfrak{H}$, $m(C) > 0$, $C \subseteq U \setminus U'$, whence (since $U'$ covers $\mathfrak{H}$) $C \subseteq U'$ mod $m$ contradicting $m(C) > 0$. This shows that $U \subseteq U'$ mod $m$ and by symmetry $U = U'$ mod $m$. The next lemma shows existence of measurable unions.

**Theorem 2.1.9.** *(Exhaustion lemma) Let $(X, \mathfrak{B}, m)$ be a probability space and let $\mathfrak{H} \subseteq \mathfrak{B}$ be hereditary, then $\exists\ A_1, A_2, \ldots \in \mathfrak{H}$ disjoint such that $U(\mathfrak{H}) = \bigcup_{n=1}^{\infty} A_n$ is a measurable union of $\mathfrak{H}$.*

*Proof.* Let

$$\epsilon_1 := \sup\{m(A) : A \in \mathfrak{H}\},$$

choose $A_1 \in \mathfrak{H}$ such that $m(A_1) \geq \frac{\epsilon_1}{2}$, and let

$$\epsilon_2 := \sup\{m(A) : A \in \mathfrak{H}, A \cap A_1 = \emptyset\},$$

choose $A_2 \in \mathfrak{H}$ such that $A_2 \cap A_1 = \emptyset$ and $m(A_2) \geq \frac{\epsilon_2}{2}$. Continuing the process, we obtain a sequence of disjoint $\{A_n\}_{n=1}^{\infty} \subseteq \mathfrak{H}$ and decreasing $\{\epsilon_k\}_k$ such that

$$\epsilon_n := \sup\{m(A) : A \in \mathfrak{H}, A \cap A_k = \emptyset\ \ \forall k < n\}, \quad m(A_n) \geq \frac{\epsilon_n}{2}.$$

Clearly, $\sum_{n=1}^{\infty} \epsilon_n \leq 2 \sum_{n=1}^{\infty} m(A_n) \leq 2$, whence $\epsilon_n \to \infty$. We claim that $U := \bigcup_{n=1}^{\infty} A_n = X$ is a measurable union of $\mathfrak{H}$. Evidently $\mathfrak{H}$ saturates $U$. To see that $U$ covers $\mathfrak{H}$ assume otherwise, then $\exists A \in \mathfrak{H}$, $m(A) > 0$ such that $A \cap A_n = \emptyset\ \forall n \geq 1$ whence $m(A) \leq \epsilon_n \to 0$ contradicting $m(A) > 0$. □

One can extend the previous result to the case when $(X, \mathfrak{B}, m)$ is $\sigma$-finite. In that case $X = \bigcup_{j=1}^{\infty} B_j$ for some measurable finite-measure sets $B_j$ and one can apply Theorem 2.1.9 to $\mathfrak{H} \cap B_j$, and then make a union over all $B_j$.

**Remark 2.1.10.** The Exhaustion Lemma tells us two things: that measurable unions exists *and* that they can be expressed by countable unions. In this notes we actually need only the first part but also the second is important.

We are ready to define the *conservative* and *dissipative part* of a non singular transformation.

**Definition 2.1.11.** The *dissipative part* of the non singular transformation $T$ is $\mathfrak{D}(T) := U(\mathcal{W}(T))$, i.e. the measurable union of the collection of wondering sets for $T$. The non-singular transformation $T$ is called (totally) *dissipative* if $\mathfrak{D}(T) = X \mod m$.

Since $T^{-1}\mathcal{W}(T) \subseteq \mathcal{W}(T)$, we have that $T^{-1}\mathfrak{D} \subseteq \mathfrak{D} \mod m$.

**Definition 2.1.12.** The set $\mathfrak{C}(T) := X \setminus \mathfrak{D}(T)$ is called the *conservative part* of $T$. The non-singular transformation $T$ is called (totally) *conservative* if $\mathfrak{C}(T) = X \mod m$.

**Definition 2.1.13.** The *Hopf decomposition* of $T$ is the partition of $X$ $\{\mathfrak{C}(T), \mathfrak{D}(T)\}$.

Let us see a first corollary of the Halmos' Recurrence Theorem.

**Corollary 2.1.14.** *If $T$ is a non-singular transformation, then*

$$\mathfrak{C}(T^n) = \mathfrak{C}(T) \mod m, \ \forall n \geq 1.$$

*Proof.* We prove that $\mathfrak{D}(T^n) = \mathfrak{D}(T)$. Since $\mathcal{W}(T) \subseteq \mathcal{W}(T^n)$ it is clear that $\mathfrak{D}(T) \subseteq \mathfrak{D}(T^n)$. On the other hand, if $W \in \mathcal{W}(T^n)_+$, then $T^{-j}W \in \mathcal{W}(T^n)_+$ and

$$\sum_{k=1}^{\infty} 1_W \circ T^k = \sum_{j=1}^{n} \sum_{k=0}^{\infty} 1_{T^{-j}W} \circ T^{nk} \leq n \text{ a.e. on } W.$$

By Halmos' Recurrence Theorem, there exists $W' \in \mathcal{W}(T)_+ \cap W$. Thus $m(W \smallsetminus \mathfrak{D}(T)) = 0$ $\forall W \in \mathcal{W}(T^n)$, whence

$$\mathfrak{D}(T^n) \subseteq \mathfrak{D}(T) \qquad \mod m.$$

□

Now, we have defined both conservative and recurrent transformations. In the next theorem and subsequent corollary we prove that these two definitions are actually equivalent. Let $B(z, \epsilon)$ be the ball at $z$ with radius $\epsilon$.

**Theorem 2.1.15.** *(Poincaré's Recurrence Theorem) Suppose that $T : X \to X$ is a conservative, non-singular transformation of $(X, \mathfrak{B}, m)$. If $(Z, d)$ is a separable metric space, and $f : X \to Z$ is a measurable map, then*

$$\liminf_{n\to\infty} d(f(x), f(T^n x)) = 0 \quad \textit{for a.e. } x \in X.$$

*Proof.* Let $Z_1 = \{z \in Z : m(f^{-1}(B(z, \epsilon))) > 0 \ \forall \epsilon > 0\}$, a closed subset of $Z$. Since $(Z, d)$ is separable, $\exists \{B(x_n, \epsilon_n)\}_n \subseteq Z_1^c$ such that $Z_1^c = \bigcup_n B(x_n, \epsilon_n)$ whence $m(X \smallsetminus f^{-1}(Z_1)) = 0$. We show that $\liminf_{n\to\infty} d(f(x), f(T^n x)) = 0$ a.e. for $x \in X_1 := f^{-1}(Z_1)$. By the Halmos' Recurrence Theorem, for $z \in Z_1$ and $\epsilon > 0$,

$$\sum_{k=1}^{\infty} 1_{f^{-1}(B(z,\epsilon))} \circ T^k = \infty \text{ a.e. on } f^{-1}(B(z, \epsilon)),$$

whence

$$\liminf_{n\to\infty} d(f, f \circ T^n) < 2\epsilon \text{ a.e. on } f^{-1}B(z, \epsilon), \forall z \in Z_1, \epsilon > 0.$$

Fix $\epsilon > 0$. By separability, there exists a sequence $\{z_n : z_n \in Z_1, n \in \mathbb{N}^+\}$ such that $Z_1 = \bigcup_{n=1}^{\infty} B(z_n, \epsilon)$, whence

$$\liminf_{n\to\infty} d(f, f \circ T^n) < 2\epsilon \text{ a.e. on } f^{-1}Z_1 = X_1.$$

□

We take inspiration form the preceding theorem to define yet another kind of recurrence.

**Definition 2.1.16.** If $(Z, d)$ is a separable metric space, we call a non singular transformation $T$ on $(X, \mathfrak{B}, m)$ *Z-recurrent* if $\liminf_{n\to\infty} d(f, f \circ T^n) = 0$ a.e. whenever $f : X \to Z$ is measurable.

**Theorem 2.1.17.** *Let $T$ be a non singular transformation of $(X, \mathfrak{B}, m)$. The following are equivalent.*

1. *$T$ is $Z$-recurrent for some separable metric space $(Z, d)$ containing at least two points.*
2. *$T$ is $Z$-recurrent for every separable metric space $(Z, d)$.*
3. *$T$ is conservative.*

*Proof.* By Theorem 2.1.15, (3) $\Rightarrow$ (2) and clearly (2) $\Rightarrow$ (1). We are left to prove that (1) $\Rightarrow$ (3). Let $A \in \mathfrak{B}_+$ and take $z_1, z_2 \in Z$ such that $d(z_1, z_2) \geq c$ for some fixed $c > 0$. Then define $f_A : X \to Z$ by

$$f_A(x) = \begin{cases} z_1, & \text{if } x \in A \\ z_2, & \text{if } x \in A^c \end{cases}$$

Clearly $f_A$ is measurable. Consider the ball $B(z_1, \epsilon)$ centered at $z_1$ with radius $\epsilon$. If we choose $\epsilon < c$, we have

$$\liminf_n d(f_A \circ T^n(x), f_A(x)) = 0 \text{ a.e. on } f^{-1}(B(z_1, \epsilon)) \Longrightarrow \sum_{k=1}^{\infty} 1_A \circ T^k(x) = \infty.$$

Therefore, the thesis follows from Halmos' Recurrence Theorem. □

**Corollary 2.1.18.** *Every recurrent transformation in the sense of Definition 2.1.4 is recurrent in the sense of Definition 2.1.3.*

*Proof.* Indeed the Definition 2.1.4 of recurrence it is just $\mathbb{R}$-recurrence in the language of Definition 2.1.16. By Theorem 2.1.17 this implies conservativity of $T$ and therefore by Halmos' Recurrence theorem Definition 2.1.3 is satisfied. □

## 2.2 Conditions for conservativity

If there exists a finite, $T$-invariant measure $q \ll m$, then clearly there can be no wondering sets with positive $q$-measure (the disjoint union of their pre-images would build up an infinite $q$-measure). Therefore, $q(\mathfrak{D}) = 0$ and $\{\frac{dq}{dm} > 0\} \subseteq \mathfrak{C} \mod m$. In particular, any probability preserving transformation is conservative. A measure-preserving transformation of a $\sigma$-finite, infinite measure space may not be conservative. For example, $x \mapsto x + 1$ is a measure preserving transformation of $\mathbb{R}$ equipped with Borel sets and Lebesgue measure, which is totally dissipative. The following propositions help to establish conservativity of measure preserving transformations of $\sigma$-finite measure spaces.

**Proposition 2.2.1.** *Suppose that $T$ is a measure preserving transformation of the $\sigma$-finite measure space $(X, \mathfrak{B}, m)$.*

1. *If $f \in L^1(m), f \geq 0$, then*

$$\left\{\sum_{k=1}^{\infty} f \circ T^k = \infty\right\} \subseteq \mathfrak{C}(T) \qquad \mod m.$$

2. *If $f \in L^1(m), f > 0$ a.e., then*

$$\left\{\sum_{k=1}^{\infty} f \circ T^k = \infty\right\} = \mathfrak{C}(T) \qquad \mod m.$$

*Proof.* We first show (1). Let $f \in L^1(m)_+$, it is sufficient to show that $\sum_{k=1}^{\infty} f \circ T^k < \infty$ a.e. on any $W \in \mathcal{W}$. To see this let $W \in \mathcal{W}$ and $n \geq 1$, then

$$\int_W \sum_{k=0}^{n} f \circ T^k dm = \sum_{k=0}^{n} \int_X 1_W f \circ T^{n-k} dm = \sum_{k=0}^{n} \int_X 1_W \circ T^k f \circ T^n dm =$$
$$= \int_X (\sum_{k=0}^{n} 1_W \circ T^k) f \circ T^n dm \leq \int_X f \circ T^n dm = \int_X f dm.$$

Therefore,

$$\int_W \sum_{k=0}^{n} f \circ T^k dm \leq \int_X f dm < \infty$$

and

$$\sum_{k=0}^{n} f \circ T^k dm < \infty \text{ a.e. on } W.$$

We now prove (2). In view of (1), for $f \in L^1(m)$, $f > 0$ a.e, it suffices to show that

$$\mathfrak{C}(T) \subseteq \{\sum_{k=1}^{\infty} f \circ T^k = \infty\} \qquad \mod m.$$

Indeed, for all $A \in \mathfrak{B}_+$, $A \subseteq \mathfrak{C}(T)$, there exists $B \subseteq A$ and $\epsilon > 0$ such that $f \geq \epsilon 1_B$ on $B$, whence

$$\sum_{k=1}^{\infty} f \circ T^k \geq \epsilon \sum_{k=1}^{\infty} 1_B \circ T^k = \infty, \text{ a.e. on } B,$$

by Halmos' Recurrence Theorem. □

The last result we present is the following theorem.

**Theorem 2.2.2.** *Suppose that $T$ is a measure preserving transformation of the $\sigma$-finite measure space $(X, \mathfrak{B}, m)$. If there exists $A \in \mathfrak{B}$, $m(A) < \infty$ such that*

$$X = \bigcup_{n=0}^{\infty} T^{-n} A \qquad \text{mod } m,$$

*then $T$ is conservative.*

*Proof.* Note that

$$X = T^{-k} X = \bigcup_{n=k}^{\infty} T^{-n} A \qquad \text{mod } m,$$

for all $k \geq 1$. Therefore, $\sum_{n=1}^{\infty} 1_A T^n = \infty$ almost everywhere. The thesis then follows from Proposition 2.2.1 with $f = 1_A$. □

# Chapter 3

# Measurable partitions, $\sigma$-algebras and disintegration

The aim of this chapter is to present Rokhlin theorem on the disintegration of measures. Our main references are [W], [CK] and Appendix B of [B]. The original proof of the Rokhlin theorem is presented in [R1] and other useful and related materials (with greater emphasis on entropy theory) is in [R2].

## 3.1 Abstract Measure Theory

We report here some concepts of abstract measure theory as presented in the very clear exposition of [CK]. There are three 'lenses' trough which we can view measure theory; we may think of it in terms of *points*, *sets* or *functions*. Suppose we have a triple $(X, \mathfrak{B}, m)$ comprising a measure space, a $\sigma$-algebra and a measure. Then we may focus on the set $X$ (and concern ourselves with points), or on the $\sigma$-algebra $\mathfrak{B}$ (and concern ourselves with sets), or on the space $L^2(X, \mathfrak{B}, m)$ (and concern ourselves with functions). All three point of view play an important role in dynamics, and various definitions and results can be given in terms of any of the three. We will see later that *the last two are completely equivalent in the greatest generality but the first requires certain additional, albeit natural, assumptions* explained in the next section. To be even more specific, it will be of particular interest to us *the correspondence between partitions of the space $X$, sub-$\sigma$-algebras of $\mathfrak{B}$ and subspaces of $L^2(X, \mathfrak{B}, m)$*.

First let us consider the set of all partitions of $X$. This is a partially ordered set, with ordering given by refinement; given two partitions $\xi$, $\eta$, we say that $\xi$ is a *refinement* of $\eta$, written $\xi \geq \eta$, if and only if every $C \in \xi$ is contained in some $D \in \eta$. In this case, we also say that $\eta$ is a *coarsening* of $\xi$. The finest partition is the partition into points, denoted $\epsilon$, while the coarsest is the trivial partition $\{X\}$, denoted $\nu$. As on any partially ordered set, we have a notion of *join (or product)* and *meet (or intersection)*, corresponding to least upper bound and greater lower bound, respectively. Given two partitions $\xi$ and $\eta$, their join is

$$\xi \vee \eta := \{C \cap D | C \in \xi, D \in \eta\}. \tag{3.1.1}$$

This is the coarsest partition which refines both $\xi$ and $\eta$, and is also sometimes referred to as the *joint partition*. For a finite or countable family $\{\xi_n\}$, the join $\xi = \bigvee_n \xi_n$ is a partition whose elements are (nonempty) intersections $\cap_n C_n$, where $C_n \in \xi_n$ for all n. One can also consider the case in which $\{\xi_\alpha\}_\alpha$ is not countable and the join is still defined (as the coarsest partition which is finer than each $\xi_\alpha$), but it has not that simple characterization anymore. The meet of $\xi$ and $\eta$ is the finest partition which coarsen both $\xi$ and $\eta$, and is denoted $\xi \wedge \eta$. In general there is no simple analogue to (3.1.1) for $\xi \wedge \eta$ (even in the countable case!). Indeed, given a finite or countable family of measurable partitions $\{\xi_m\}$, one may attempt to construct their meet $\bigwedge_m \xi_m$ as follows. For any $x, y \in X$ we put $x \sim y$ if there exists a finite sequence $A_1, ..., A_n \in \cup_m \xi_m$ such that $x \in A_1$ and $y \in A_n$ and $A_i \cap A_{i+1} \neq \emptyset$ for all $1 \leq i \leq n-1$ (of course for each $i$ the sets $A_i$ and $A_{i+1}$ must be elements of different partitions here, because their intersection in non-empty). Then $\sim$ will be an equivalent relation on $X$ and its classes naturally define a partition of $X$; we call it $\tilde{\wedge}_m \xi_m$. The partition $\tilde{\wedge}_m \xi_m$ corresponds, 'logically', to the meets $\wedge_m \xi_m$, and in many cases $\wedge_m \xi_m = \tilde{\wedge}_m \xi_m \mod 0$. But in some cases $\tilde{\wedge}_m \xi_m$ may be quite different from $\wedge_m \xi_m$ (in fact, $\tilde{\wedge}_m \xi_m$ may not even be measurable), so it is not safe to use it in lieu of $\wedge_m \xi_m$.

**Example 1.** Let $X = [0, 1]$ and $\mu$ the Lebesgue measure on $X$. Let $\xi_1 = \{[0, 0.5], (0, 5, 1]\}$ and $\xi_2 = \{\{0, 1\}, \{x\}\}_{x \in X \setminus \{0,1\}}$ ($\xi_2$ is a partition of $X$ into the set $\{0, 1\}$ and all one-point subsets of $X \setminus \{0, 1\}$). As one can verify[1]

$$\xi_1 \wedge \xi_2 = \xi_1 \quad \text{but} \quad \xi_1 \tilde{\wedge} \xi_2 = \nu.$$

Anyway we mention that this problem can be fixed when we are in a Lebesgue space. In that case one can prove that there exists a full measure set $D \subset X$ such that if we denote by $\xi_D$ the partition $\xi$ restricted to the set $D$ (i.e. $\xi_D = \{C \cap D | C \in \xi\}$ it holds

$$\tilde{\wedge}_m \xi_{m,D} = \wedge_m \xi_{m,D},$$

whenever $\tilde{\wedge}_m \xi_{m,D}$ is measurable (the interested reader may look at the Appendix B of [CM]).

Given a partition $\xi$, we may consider the collection of all measurable subsets $A \subset X$, which are unions of elements of $\xi$; this collection forms a sub-$\sigma$-algebra of $\mathfrak{B}$ which we

[1] Note that the equality for partitions below is to be intended mod 0, a concept that we will clarify better later.

will denote by $\mathfrak{F}(\xi)$. We will see later that *this correspondence is far from injective*; for example, certain partitions whose elements are countable sets are associated with the trivial $\sigma$-algebra.
Similarly, we can consider the collection of all square integrable functions which are constant on elements of $\xi$; the collection of all equivalence classes of such functions forms the subspace $L^2(X, \mathfrak{F}(\xi), \mu) \subset L^2(X, \mathfrak{B}, \mu)$.

**Example 2.** Let $X = [0,1] \times [0,1]$ be the unit square with Lebesgue measure $Leb_2$ and $\mathfrak{B}$ the Borel $\sigma$-algebra. Let $\xi = \Big\{\{x\} \times [0,1] | x \in [0,1]\Big\}$ be the partition into vertical lines; then $\mathfrak{F}(\xi)$ is the sub-$\sigma$-algebra consisting of all sets of the form $E \times [0,1]$, where $E \subseteq [0,1]$ is Borel, and $L^2(X, \mathfrak{F}(\xi), Leb_2)$ is the space of all square-integrable functions which depend only on the $x$-coordinate. The latter is canonically isomorphic to the space of square-integrable functions on the unit interval.

The map $\xi \to \mathfrak{F}(\xi)$ is a morphisms of partially ordered sets; it is natural to ask whether this morphism in injective (and hence invertible) on a certain class of partitions. First, however, we turn to the question of classifying measure spaces and hence the associated class of $\sigma$-algebras and partitions, since the end result turns out to be relatively simple. It turns out that such a classification will also be crucial for the problem of correspondence between partition and sub-$\sigma$-algebras.

## 3.2 Lebesgue spaces

It is somewhat serendipitous that although one may consider many different measure spaces $(X, \mathfrak{B}, \mu)$, which on the face of it are quite different from each other, all of the examples in which we will be interested actually fall into a relatively simple classification. To elucidate this statement, let us consider the two fundamental examples of measure spaces. The simplest sort of measure space is an *atomic space*, in which $X$ is a finite or countable space. At the other hand of the spectrum stand the *non-atomic spaces*, in which every positive measure set can be decomposed into two subsets of smaller positive measure. The easiest example of such a space is the interval $[0,1]$ with Lebesgue measure. *In fact, up to isomorphism and setting aside examples which are for our purpose pathological, this is the only example of such a space.*

### 3.2.1 Isomorphism and conjugacy

What does it mean for measure space to be isomorphic? The most immediate (and essentially correct) idea is to require existence of a bijection between the spaces which carries measurable sets into measurable sets both ways and preserves the measure. Obviously such a bijection must be in accordance with one of the fundamental principles of measure theory: disregarding null-measure sets.

**Definition 3.2.1.** (Measure space isomorphism) The probability spaces $(X_1, \mathfrak{B}_1, m_1)$ and $(X_2, \mathfrak{B}_2, m_2)$ are said to be *isomorphic* if there exists $M_1 \in \mathfrak{B}_1$ and $M_2 \in \mathfrak{B}_2$ with $m_1(M_1) = 1$ and $m_2(M_2) = 1$ and an invertible measure-preserving transformation $\phi : M_1 \to M_2$. (The space $M_i$ is assumed to be equipped with the $\sigma$-algebra $M_i \cap \mathfrak{B}_i = \{M_i \cap B | B \in \mathfrak{B}_i\}$ and the restriction of the measure $m_i$ to this $\sigma$-algebra.)

This definition cares about *points* because we ask $\phi$ to be an isomorphisms. If we decide to focus on *sets*, the most natural way of define an 'isomorphism' disregarding sets of zero measure is trough what is called $\sigma$-algebras isomorphisms or *conjugacy*. This requires that we define first measure algebras.

**Definition 3.2.2.** Let $(X, \mathfrak{B}, m)$ be a probability space. Define an equivalence relation on $\mathfrak{B}$ by saying that $A$ and $B$ are equivalent $(A \sim B)$ iff $m(A \triangle B) = 0$. Let $\tilde{\mathfrak{B}}$ denote the collection of equivalence classes. Then $\tilde{\mathfrak{B}}$ is a Boolean $\sigma$-algebra under the operation of complementation, union and intersection inherited from $\mathfrak{B}$. The measure $m$ induces a measure $\tilde{m}$ on $\tilde{\mathfrak{B}}$ by $\tilde{m}(\tilde{B}) = m(B)$. (Here $\tilde{B}$ is the equivalence class to which $B$ belongs). The pair $(\tilde{\mathfrak{B}}, \tilde{m})$ is called a measure algebra.

**Definition 3.2.3.** Let $(X_1, \mathfrak{B}_1, m_1)$ and $(X_2, \mathfrak{B}_2, m_2)$ be probability spaces with measure algebras $(\tilde{\mathfrak{B}}_1, \tilde{m}_1)$ and $(\tilde{\mathfrak{B}}_2, \tilde{m}_2)$. The measure algebra are isomorphic if there is a bijection $\Phi : \tilde{\mathfrak{B}}_2 \to \tilde{\mathfrak{B}}_1$ which preserves complements, countable unions and intersections and satisfies $\tilde{m}_1(\Phi \tilde{B}) = \tilde{m}_2(\tilde{B})$ for any $\tilde{B} \in \tilde{\mathfrak{B}}_2$. The probability spaces are said to be *conjugate* if their measure algebras are isomorphic.

Conjugacy of measure spaces is weaker than isomorphism because if $(X_1, \mathfrak{B}_1, m_1)$ and $(X_2, \mathfrak{B}_2, m_2)$ are isomorphic as in Definition 3.2.1 then they are conjugate via $\Phi : \tilde{\mathfrak{B}}_2 \to \tilde{\mathfrak{B}}_1$ defined by $\Phi(\tilde{B}) := (\phi^{-1}(M_2 \cap B))^{\sim}$. On the other hand, it is easy to give examples of conjugate measure spaces which are not isomorphic. Let $(X_1, \mathfrak{B}_1, m_1)$ be a space of one point and $(X_2, \mathfrak{B}_2, m_2)$ be a space of two points and $\mathfrak{B}_2 = \{\emptyset, X_2\}$. The two spaces are conjugate but they are not isomorphic because a set of zero measure cannot be omitted from $X_2$ so that the remaining set is mapped bijectively with $X_1$. *The main reason this example works is that* $\mathfrak{B}_2$ *does not separate the points of* $X_2$. What is nice about Lebesgue spaces is that they are designed in such a way that this never happens and everything works so that isomorphism and conjugacy are equivalent. So, actually, in these spaces there is in principle no conceptual gap between a point-description and a set-description of measure theory. What is even nicer about Lebesgue spaces is that they are isomorphic (with the stronger Definition 3.2.1 and so also with the weaker one 3.2.3), with the measure space $([0,1], \mathfrak{L}, Leb_1)$, where $\mathfrak{L}$ are the Borel subsets[2] of $[0,1]$ and $Leb_1$ is the Lebesgue measure. So, to sum up:

1. In Lebesgue spaces isomorphism and conjugacy are equivalent;

[2]Actually the Lebesgue subsets, i.e. the completion of the Borel $\sigma$-algebra, because we assume always to deal with complete $\sigma$-algebras, see the next section.

2. Lebesgue spaces are isomorphic to $([0,1], \mathfrak{L}, Leb_1)$.

In some books such as [W] Lebesgue spaces are defined precisely by saying that they are isomorphic to $([0,1], \mathfrak{L}, Leb_1)$ (with a possible addition of countably many atoms). We follow another route: in the next section we will present a couple of theorems that basically say that if the measure space has some good properties then it is more and more similar to $([0,1], \mathfrak{L}, Leb_1)$. We then put in the defintion of Lebesgue space exactly the conditions one needs to make a generic measure space isomorphic to $([0,1], \mathfrak{L}, Leb_1)$. In doing so, we will find out that these conditions are precisely the ones that make isomorphisms equivalent to conjugacy.

### 3.2.2 Defining Lebesgue spaces

Let us introduce a metric $d_\mu$ on the measure algebra $(\tilde{\mathfrak{B}}, \tilde{m})$ by the formula

$$d_\mu(\tilde{A}, \tilde{B}) = \mu(A \triangle B),$$

where $A$ and $B$ are some representative elements of the equivalence classes $\tilde{A}$ and $\tilde{B}$, respectively. $(\tilde{\mathfrak{B}}, d_\mu)$ is a metric space.

**Definition 3.2.4.** We say that the $\sigma$-algebra $\mathfrak{B}$ is *separable* if $(\tilde{\mathfrak{B}}, d_\mu)$ is separable as a metric space. That is, $\mathfrak{B}$ is separable iff there exists a countable collection of equivalence classes of set $\{\tilde{A}_n\}_{n\in\mathbb{N}}$ which is dense in $\tilde{\mathfrak{B}}$ with the metric $d_\mu$. If one consider representative elements $A$ of each $\tilde{A}$ then $d_\mu$ is not anymore a metric but becomes a pseudo-metric (two different sets can have zero distance). In this case $\mathfrak{B}$ is separable iff there exists a countable collection of sets $\{A_n\}_{n\in\mathbb{N}}$ which is dense in $\mathfrak{B}$ with the pseudo-metric $d_\mu$.

**Definition 3.2.5.** The *completion* of $\mathfrak{B}$ is the $\sigma$-algebra generated by $\mathfrak{B}$ together with all subsets of null sets (one can prove that this collection of sets is indeed a $\sigma$-algebra, see e.g. [Sc]). We say that two $\sigma$-algebra are equivalent $\mod 0$ if they have the same completion.

From now on we will consider only complete $\sigma$-algebras, unless otherwise specified, and when we will say that two $\sigma$-algebras are equal we may forget to say that they are equal $\mod 0$.
Recall that if $\mathcal{A} \subseteq \mathfrak{B}$ is any collection of sets, then $\sigma(\mathcal{A})$, the $\sigma$-algebra generated by $\mathcal{A}$, is the smallest $\sigma$-algebra that contains $\mathcal{A}$.

**Proposition 3.2.6.** *Let $(X, \mathfrak{B}, m)$ be a probability space. If $\mathfrak{B}$ is equivalent mod zero to a $\sigma$-algebra generated by a countable collection of sets, then $\mathfrak{B}$ is separable.*

*Proof.* Let $\mathcal{A} = \{A_n\}_{n\in\mathbb{N}}$ be the countable collection of sets that generates $\mathfrak{B}$. Let $\mathcal{A}^* = \{A_k^*\}_{k\in\mathbb{N}}$ be the set of all finite unions of elements of $\mathcal{A}$, which is countable. Let us denote $\bar{\mathcal{A}}^*$ the closure of $\mathcal{A}^*$ un the $d_\mu$ metric. Clearly $\bar{\mathcal{A}}^*$ contains $\mathcal{A}$. It can also be shown that $\bar{\mathcal{A}}*$ is closed under intersection and countable union. This shows that $\bar{\mathcal{A}}^*$ contains $\mathfrak{B}$ and hence $\mathcal{A}^*$ is dense in $\mathcal{B}$. □

**Corollary 3.2.7.** *Let $X$ be a separable metric space, $\mathfrak{B}$ its the Borel $\sigma$-algebra and $\mu$ a probability measure. Then $\mathfrak{B}$ is separable as a $\sigma$-algebra.*

**Theorem 3.2.8.** *Let $(X, \mathfrak{B}, m)$ be a probability space. Let also $(\tilde{\mathfrak{B}}, \tilde{m})$ be a separable measure algebra (complete) with no atoms. Then $(\tilde{\mathfrak{B}}, \tilde{m})$ is isomorphic to the $\sigma$-algebra of Lebesgue sets of the unit interval. In other words, the probability space $(X, \mathfrak{B}, m)$ is conjugated to $([0,1], \mathfrak{L}, Leb_1)$.*

*Proof.* (Sketch) Given a countable collection of sets $\{A_n\} \in \mathfrak{B}$, write $A_n^0 := A_n$ and $A_n^1 := X \setminus A_n$, and let $\xi_n = \bigvee_{i=1}^n \{A_n^0, A_n^1\}$. This is an increasing sequence of partitions of $X$, and because $\mathfrak{B}$ is separable, we may take the sets $A_n$ to be such that $\hat{\mathfrak{B}} := \bigcup_{n\geq 1} \mathfrak{B}(\xi_n)$ is dense in $\mathfrak{B}$. Moreover $\xi_n$ has $\leq 2^n$ elements that can be indexed by words $w = w_1...w_n \in \{0,1\}^n$ as follows:

$$A_w = A_1^{w_1} \cap A_2^{w_2} \cap ... \cap A_n^{w_n}.$$

Write $\mathfrak{L}$ for the $\sigma$-algebra of Lebesgue sets of the unit interval. We can define a map $\rho : \hat{\mathfrak{B}} \to \mathfrak{L}$ as follows.

1. For a fixed $n$, order the elements of $\xi_n$ lexicographically: for example, with $n = 3$ we have
$$A_{000} \leq A_{001} \leq A_{010} \leq A_{011} \leq ... \leq A_{111}.$$

2. Identify these sets with sub-intervals of $[0,1]$ with the same measure and the same order: thus
$$\rho(A_{000}) = [0, \mu(A_{000})),$$
$$\rho(A_{001}) = [\mu(A_{000}), \mu(A_{000}) + \mu(A_{001})),$$
and so on. In general we have
$$\rho(A_w) = \Big[\sum_{v<w} \mu(A_v), \sum_{v\leq w} \mu(A_v)\Big),$$
where the sums are over words $v$ of the same length as $w$. Note that the image is empty whenever the sums are equal, which happens exactly when $\mu(A_w) = 0$.

It is clear form the construction that $\rho$ respects countable union and complements, i.e.

$$\rho\Big(\bigcup_{n=1}^{\infty} E_i\Big) = \bigcup_{n=1}^{\infty} \rho(E_i),$$

$$\rho(X \setminus E) = [0,1] \setminus \rho(E),$$

whenever the union is an element of $\hat{\mathfrak{B}}$. Furthermore, $\rho$ is an isometry with respect to the metrics $d_\mu$ and $d_{Leb_1}$, and so it can be extended from the dense set $\hat{\mathfrak{B}}$ to all of $\mathfrak{B}$. The identity $Leb_1(\rho(E)) = \mu(E)$ holds on $\hat{\mathfrak{B}}$ and extend to $\mathfrak{B}$ by continuity. Finally, because $\mathfrak{B}$ is non-atomic, the measure of all the sets in $\xi_n$, goes to 0, and so the length of their images under $\rho$ goes to 0 as well, which shows that $\rho(\mathfrak{B}) = \mathfrak{L}$. □

The proof of Theorem 3.2.8 describes the construction of a $\sigma$-algebra isomorphism (or conjugacy). At a first glance, it may appear that this creates an isomorphism between the measure spaces themselves, but in fact, the proof as it stands may not produce such an isomorphism. The problem is the potential presence of "holes" in the space $X$ distributed among its points in a non-measurable way. In order to give a classification result for measure spaces themselves, we need one further condition in addition to separability which prevents the appearance of "holes".
Let $\xi_1 = \{C_1, ..., C_n\}$ be a finite partition of $X$ into measurable sets, and let $\mathfrak{B}_1 = \mathfrak{F}(\xi_1)$ be the $\sigma$-algebra which contains all union of elements of $\xi_1$, so that $\mathfrak{B}_1$ contains $2^n$ sets. Partitioning each $C_i$ into $C_{i,1}, ..., C_{i,k_i}$, we obtain a finer partition $\xi_2$ and a larger $\sigma$-algebra $\mathfrak{B}_2 = \mathfrak{F}(\xi_2)$ whose elements are unions of none, some, or all of the $C_{i,j}$. Iterating this procedure, we have a sequence of partitions

$$\xi_1 < \xi_2 < ..., \tag{3.2.1}$$

each of which is a refinement of the previous partition, and a sequence of $\sigma$-algebras

$$\mathfrak{B}_1 \subset \mathfrak{B}_2 \subset ....$$

This is obviously reminiscent of the construction from the proof of Theorem 3.2.8. An even more useful image to keep in mind here is the standard picture of the construction of a Cantor set, in which the unit interval is first divided into two pieces, then four, then eight, and so on - these "cylinders" (to use the terminology from symbolic dynamics) are the various sets $C_i, C_{i,j}, C_{i,j,k}$, etc.
We may consider the "limit" of the sequence (3.2.1):

$$\xi = \bigvee_{n=1}^{\infty} \xi_n.$$

Each element of $\xi$ corresponds to a "funnel"

$$C_{i_1} \subset C_{i_1,i_2} \subset C_{i_1,i_2,i_3} \subset ...$$

of decreasing subsets within the sequence of partitions; the intersection of all the sets in such a funnel is an element of $\xi$.

**Definition 3.2.9.** The sequence (3.2.1) is called *basis* for $(X, \mathfrak{B})$ if it generates *both* the $\sigma$-algebra $\mathfrak{B}$ *and* the space $X$, as follows:

1. the associated $\sigma$-algebras $\mathfrak{B}_n := \mathfrak{F}(\xi_n)$ have the property that $\bigcup_{n\geq 1} \mathfrak{B}_n$ generates $\mathfrak{B}$;

2. it generates the space $X$; that is every "funnel" $C_{i_1} \subset C_{i_1,i_2} \subset C_{i_1,i_2,i_3} \subset ...$ has intersection containing at most one point. Equivalently, any two point $x$ and $y$ are separated by some partition $\xi_n$, and so $\xi := \bigvee_{n=1}^{\infty} \xi_n = \epsilon$, the partition into points.

Note that the existence of an increasing sequence of finite or countable partitions satisfying (1) is equivalent to the separability of the $\sigma$-algebra.
It is often convenient to choose a sequence $\xi_n$ such that at each stage, each cylinder set $C$ is partitioned into exactly two smaller sets. This gives a one to one correspondence between sequences in $\Sigma_2^+ := \{0,1\}^{\mathbb{N}}$ and "funnels". Such a "funnel" corresponds to some element of $\mathfrak{B}$ which is either a singleton or empty, we have associated to each Borel subset of $\Sigma_2^+$ an element of $\mathfrak{B}$, and so $\mu$ yields a measure on $\Sigma_2^+$. Thus we have a notion of 'almost all funnels'.

**Definition 3.2.10.** We say that the basis is *complete* if almost every funnel contains exactly one point[3]. That is, the set of funnels whose intersection is empty should be measurable and should have measure zero. Equivalently, a basis defines a map from $X$ to $\Sigma_2^+$ which takes each point to the "funnel" containing it; the basis is complete if the image of this map has full measure.

The existence of a complete basis is the final invariant needed to classify "nice" measure spaces.

**Theorem 3.2.11.** *If $(X, \mathfrak{B}, \mu)$ is separable, non-atomic, and posses a complete basis, then it is isomorphic to Lebesgue measure on the unit interval.*

*Proof.* Full details can be found in [R1]; here we describe the main idea, which is that with completeness assumption, the argument of Theorem 3.2.8 indeed gives an isomorphism of measure spaces. Using the notation from that proof, every infinite intersection $\bigcap_{n\geq 1} A_{x_1,\dots,x_n}$ corresponds to a *single* point in the interval, namely

$$\lim_{n\to\infty} \sum_{v<x_1\dots x_n} \mu(A_v).$$

[3] This is not to be confused with the notion of completeness for $\sigma$-algebras.

With the exception of a countable set (the endpoints of basic intervals of various ranks), every point is the image of at most one "funnel". Completeness guarantees that the correspondence is indeed a bijection between sets of full measure. Measurability follows from the fact that images of the sets from a basis are finite unions of intervals. □

In fact, all the measure spaces of interest to us are separable and complete, as the following propositions shows.

**Theorem 3.2.12.** *Let $(X, \mathfrak{B}, \nu)$ be probability space, with $X$ separable metric space, $\mathfrak{B}$ the $\sigma$-algebra of Borel sets and $\nu$ probability measure. Suppose in addition that $X$ is complete as a metric space. Then $(X, \mathfrak{B}, \nu)$ has a complete basis.*

*Proof.* (Sketch) Fix $x \in X$ and let $S_r = \partial B(x, r)$ be the boundary of the ball or radius $r$ centered at $x$. Then for any $x$ at most countably many of the $S_r$ have positive measure. Indeed, if not one would have uncountably many disjoint sets of positive measure which is absurd since $\nu$ is a probability measure. Since $X$ is separable as a metric space there exists a dense countable set $\{x_n\}$. Consider the balls $\{B(x_n, r_m)\}_{n,m\in\mathbb{N}}$, one can choose the values of $r_m$ such that the boundaries $\partial B(x_n, r_m)$ have all zero measure and $r_m \to 0$ as $m \to \infty$. One has that the collection $\{B(x_n, r_m)\}_{n,m}$ forms a complete basis. Indeed, any $x \in X$ can be identified with the 'funnel' $\{B(x_n, r_n)\}$ for some sequence $x_n \to x$ and $r_n \to 0$ for $n \to \infty$. This map between funnels and points is a bijection except for those points which lie in some boundary $\partial B(x_n, r_m)$ but the set of all such points have zero measure since is a union of countably many sets of zero measure. □

**Definition 3.2.13.** A separable measure space $(X, \mathfrak{B}, m)$ with a complete basis is called a *Lebesgue space*[4].

In light of this definition we can rephrase Theorem 3.2.12 as the result that every separable complete metric space equipped with a Borel probability measure is a Lebesgue space. By Theorem 3.2.11, every Lebesgue space is isomorphic to the union of unit interval with at most countably many atoms. It is also worth noting that any separable measure space admits a completion, just as in the case for metric spaces. The procedure is quite simple; take a basis for $X$ which is not complete, and add to $X$ one point corresponding to each empty "funnel". Thus we need not concern ourselves with non-complete spaces.

We have elucidate the promised difference between the language of sets and that of points: separability is sufficient for the first to lead to the standard model, while for the second, completeness is also needed. This distinction is important theoretically; in particular, it allows us to separate results which hold for arbitrary separable measure spaces (such as ergodic theorems) from those which hold in Lebesgue spaces (such as von Neumann's isomorphism theorem for dynamical systems with pure point spectrum).

[4] Actually, separability of the measure space is implicit once we assume the existence of a basis.

However, non-Lebesgue measure spaces are at least as pathological for "normal mathematics" as non-measurable sets or sets of cardinality higher than continuum. As a consequence of Theorem 3.2.12, the measure spaces which arise in conjunction with dynamics are all Lebesgue, so for now on we will restrict our attention to them.

## 3.3 Partitions and $\sigma$-algebras

We have seen that to each partition one can associate the sub $\sigma$-algebra $\mathfrak{F}(\xi)$ just by taking measurable unions of elements of $\xi$. It is natural to ask whether this operation is 1-1. The answer is partially affirmative if we consider classes of equivalent mod 0 partitions.

**Definition 3.3.1.** Two partitions $\xi$ and $\eta$ of $X$ are *equivalent mod zero* if there exists a set $E \subset X$ of full measure such that

$$\Big\{C \cap E | C \in \xi\Big\} = \Big\{D \cap E | D \in \eta\Big\},$$

in which case we write $\xi = \eta \mod 0$.

As for $\sigma$-algebra we will always consider classes of equal mod 0 partitions and we may forget to specify the mod 0 from now on.

**Theorem 3.3.2.** *Given a separable measure space $(X, \mathfrak{B}, m)$ and a sub-$\sigma$-algebra $\mathfrak{A} \subset \mathfrak{B}$, there exists a partition $\xi$ of $X$ into measurable sets such that $\mathfrak{A}$ and $\mathfrak{F}(\xi)$ are equivalent mod zero.*

*Proof.* Without loss of generality, we may assume that $\mu$ is non-atomic (if $\mathfrak{A}$ contains any atom, these can be taken as elements of $\xi$, and there can be only countably many disjoint atoms). Since $(X, \mathfrak{B}, \mu)$ is separable, so is $(X, \mathfrak{A}, \mu)$. (The metric space $(\mathfrak{A}, d_\mu)$ is a subspace of $(\mathfrak{B}, d_\mu)$). In particular, we may take a basis $\{\xi_n\}_{n\in\mathbb{N}^+}$ for $\mathfrak{A}$ and define $\xi = \bigvee_{n=1}^{\infty} \xi_n$. One has that $\mathfrak{A} = \mathfrak{F}(\xi)$. This follows from the fact that $\{\xi_n\}$ is a basis for $\mathfrak{A}$ and so $\mathfrak{A}$ is generated by the collection of sets $\bigcup_{n\geq 1} \mathfrak{F}(\xi_n)$. But $\mathfrak{F}(\xi)$ contains this collection and is a $\sigma$-algebra by definition. Therefore $\mathfrak{F}(\xi) \supseteq \mathfrak{A}$. The opposite inclusion follows form the fact that $\mathfrak{A} \supseteq \mathfrak{F}(\xi_n)\ \forall n \in \mathbb{N}^+$ and $\mathfrak{F}(\bigvee_{n=1} \xi_n)$ is the smallest $\sigma$-algebra with that property. Therefore, $\mathfrak{A} \supseteq \mathfrak{F}(\xi)$ and equality is proved. □

We will denote the partition $\xi = \bigvee_{n=1}^{\infty} \xi_n$ constructed in Theorem 3.3.2 by $\Xi(\mathfrak{A})$.

**Definition 3.3.3.** Given a sub-$\sigma$-algebra $\mathfrak{A}$ of a separable measure space $(X, \mathfrak{B}, m)$, we define *the associated measurable partition* by this operation. Take a countable dense set $\{A_n\}_{n\in\mathbb{N}}$ of $\mathfrak{A}$ and define $\xi_n = \{A_n, X \setminus A_n\}$. Then $\Xi(\mathfrak{A}) = \bigvee_{n=1}^{\infty} \xi_n$.

Analogously to what we have done before, the elements of $\Xi(\mathfrak{A})$ may be described explicitly as follows: without loss of generality, assume that $A_n^0$ and $A_n^1 = X \setminus A_n^0$ are such that each $\xi_n$ has the form $\bigvee_{k=1}^n \{A_k^0, A_k^1\}$, and given $w \in \{0,1\}^{\mathbb{N}}$, let $A_w = \bigcap_{n\in\mathbb{N}} A_n^{w_n}$. Note that unlike in the proofs of Theorems 3.2.8 and 3.2.11 the $A_w$ are not only empty sets ore singletons; indeed, some intersections $A_w$ *must* contain more than one point unless $\Xi(\mathfrak{A}) = \epsilon$.

## 3.4 Measurable partitions

We now have a natural way to go from a partition $\xi$ to a $\sigma$-algebra $\mathfrak{F}(\xi) \subset \mathfrak{B}$, and form a $\sigma$-algebra $\mathfrak{A} \subset \mathfrak{B}$ to a partition $\Xi(\mathfrak{A})$. Theorem 3.3.2 guarantees that $\Xi(-)$ it is a one-sided inverse to $\mathfrak{F}(-)$, in the sense that $\mathfrak{F}(\Xi(\mathfrak{A})) = \mathfrak{A}$ for any $\sigma$-algebra $\mathfrak{A}$ (up to equivalence mod zero). So we may ask if it is also true that $\xi$ and $\Xi(\mathfrak{F}(\xi))$ are equivalent in some sense.
We see that since each set in $\Xi(\mathfrak{B}(\xi))$ is measurable, we should at least demand that $\xi$ not contain any non-measurable set. For example, consider the partition $\xi = \{A, B\}$, where $A \cap B = \emptyset$ and $A \cup B = X$: then if $A$ is measurable (and hence $B$ as well), we have

$$\mathfrak{B}(\xi) = \{\emptyset, A, B; X\},$$

and $\Xi(\mathfrak{B}(\xi)) = \{A, B\} = \xi$, while if $A$ is non-measurable, we have

$$\mathfrak{B}(\xi) = \{\emptyset, X\}$$

and so $\Xi(\mathfrak{B}(\xi)) = \nu$. Thus a “good” partition should only contain measurable sets; it turns out however that this is not sufficient, and that there are examples where $\Xi(\mathfrak{B}(\xi))$ is not equivalent mod zero to $\xi$, even though every set in $\xi$ is measurable.

**Example 1.** Consider the torus $\mathbb{T}^2$ with Lebesgue measure $Leb_2$, and let $\xi$ be the partition into orbits of a linear flow $\phi_t$ with irrational slope $\alpha$; that is, $\phi_t(x, y) = (x + t, y + t\alpha)$. In order to determine $\mathfrak{B}(\xi)$, we must determine which measurable sets are unions of orbits of $\phi_t$; that is, which measurable sets are invariant. Because this flow is ergodic with respect to $Leb_2$, any such set must have measure 0 or 1, and so up to sets of measure zero, $\mathfrak{B}(\xi)$ is the trivial $\sigma$-algebra. It follows that $\Xi(\mathfrak{B}(\xi))$ is the trivial partition $\nu = \{\mathbb{T}^2\}$! A discrete time version of this is the partition of the circle into orbits of an irrational rotation.

**Definition 3.4.1.** The partition $\Xi(\mathfrak{B}(\xi))$ is known as *measurable hull* of $\xi$, and will be denoted by $\mathcal{H}(\xi)$. If $\xi$ is equivalent mod zero to its measurable hull, we say that it is a *measurable partition.*

It is obvious that in general the measurable hull of $\xi$ is a coarsening of $\xi$; the definition says that if $\xi$ is non-measurable, this is a proper coarsening.

**Proposition 3.4.2.** *The measurable hull $\mathcal{H}(\xi)$ is the finest measurable partition that coarsens $\xi$. In particular, if $\eta$ is any partition with*

$$\xi \geq \eta > \mathcal{H}(\xi),$$

*then $\mathcal{H}(\eta) = \mathcal{H}(\xi)$ and hence $\eta$ is non measurable.*

*Proof.* Clearly $\mathcal{H}(\xi) \geq \mathcal{H}(\eta)$. Since $\eta > \mathcal{H}(\xi)$ we have that $\mathfrak{F}(\eta) \supseteq \mathfrak{F}(\mathcal{H}(\xi))$ and hence $\mathcal{H}(\eta) = \Xi(\mathfrak{F}(\eta)) \geq \Xi(\mathfrak{F}(\Xi(\mathfrak{F}(\xi)))) = \Xi(\mathfrak{F}(\xi)) = \mathcal{H}(\xi)$. □

To sum up, $\mathfrak{F}$ gives a map from the class of all partitions to the class of all $\sigma$-algebras, and $\Xi$ gives a map in the opposite direction, which is the one-sided inverse of $\mathfrak{F}$. We see that the set of all measurable partitions is just the image of the map $\Xi$, on which $\mathcal{H}$ acts as the identity, and on which $\mathfrak{F}$ and $\Xi$ are two-sided inverse.
If we denote by $\mathcal{O}$ the partition into orbits of some dynamical system, then $\mathcal{H}(\mathcal{O})$ is also known as the *ergodic decomposition* of the system and is denoted by $\mathcal{E}$. Example (1) shows that the orbit partition for an irrational toral flow is non-measurable; in fact this is true for *any ergodic system with more than one orbit*, since in this case $\mathfrak{F}(\mathcal{O})$ is the trivial $\sigma$-algebra, whence $\mathcal{E} = \{X\}$ is the trivial partition and $\mathcal{O} \neq \mathcal{E} = \mathcal{H}(\mathcal{O})$. This sort of phenomena is widespread in dynamical systems, for examples in the context of smooth dynamics, the partition into unstable manifolds is non-measurable whenever the entropy is positive[5].
An alternative characterization of measurability may be motivated by recalling that in the "toy" example of the partition into two subsets, the corresponding $\sigma$-algebra had four elements in the measurable case and only two in the non measurable one. In some sense, measurability of the partition corresponds to an increasing "richness" of the associated $\sigma$-algebra.

**Theorem 3.4.3.** *Let $\xi$ be a partition of a Lebesgue space $(X, \mathfrak{B}, \mu)$. $\xi$ is measurable if and only if there exists a countable collection of sets $\{A_n\}_{n\in\mathbb{N}} \subset \mathfrak{F}(\xi)$ such that for almost every pair $C_1, C_2 \in \xi$ we can find some $A_n$ that separates them in the sense that $C_1 \subseteq A_n$ and $C_2 \subseteq X \setminus A_n$.*

*Proof.* Since $\mathfrak{F}(\xi)$ is a sub-$\sigma$-algebra of the separable $\sigma$-algebra $\mathfrak{B}$, then $\mathfrak{F}(\xi)$ is separable with a dense basis $\{A_n\}_{n\in\mathbb{N}}$. Furthermore, since $\xi$ is measurable then the $\Xi(-)$ operation on $\mathfrak{F}(\xi)$ has to yield exactly $\xi$. To do so it must be that the the sets $\{A_n\}_{n\in\mathbb{N}}$ separate the elements of $\xi$ (i.e. the basis has to be complete once we regard the elements of the partition $\xi$ as the points of a new Lebesgue space). Consider now the opposite

[5]For smooth dynamical systems with singularities such as billiards, more often one is interested to the partition into *local* unstable manifold which is measurable; it is not measurable the one into *global* unstable manifolds.

implication. The key observation is the fact that the collection $\{A_n\}_{n\in\mathbb{N}}$ corresponds to a refining sequence of partitions defined by

$$\eta_k = \{A_k, X \smallsetminus A_k\}, \quad \xi_n = \bigvee_{k=1}^{n} \eta_k.$$

One can easily verify that $\{A_n\}_{n\in\mathbb{N}}$ is a basis for $\mathfrak{F}(\xi)$. Therefore, by definition $\mathcal{H}(\xi) = \bigvee_{n=1}^{\infty} \xi_n$. Since the $A_n$s separate the elements of the partition, one has that $\mathcal{H}(\xi) = \Xi(\mathfrak{F}(\xi)) \geq \xi$ and hence $\mathcal{H}(\xi) = \xi$. □

**Remark 3.4.4.** Sometimes, authors use the characterization of measurable partitions in Theorem 3.4.3 as a definition for that property. See for example Appendix B of [CM].

It may be not immediately clear what is meant by "almost every pair" in the statement of the previous theorem. Recall that the natural projection $\pi : X \to \xi$ takes $x \in X$ to the unique partition element $C \in \xi$ containing $x$. Thus the space $\xi$ is naturally endowed with a measure, i.e. the pushforward of $\mu$ under $\pi$. If we call this measure $\mu_\xi$[6] then, given a set $E \subset \xi$ (such that $\pi^{-1}(E)$ is $\mu$-measurable), we have

$$\mu_\xi(E) = \mu(\pi^{-1}(E)).$$

This gives a meaning to the notion of "almost every" and also to the one of "almost every pair". Aside from finite or countable partitions into measurable sets (which are obviously measurable), a good example of measurable partition is given by the Example (2) (in Section 3.1) of the partition of the unit square into vertical lines. In fact, this is in some sense the *only* measurable partition, just as $[0, 1]$ is, up to isomorphism, the only Lebesgue space. The following result states that a measurable partition can be decomposed into a "discrete" part, where each element has positive measure and a "continuous part", which is isomorphic to the partition of the square into lines.

**Theorem 3.4.5.** *Given a measurable partition $\xi$ of a Lebesgue space $(X, \mathfrak{B}, \mu)$ there exists a set $E \subset X$ such that*

1. *Each element of $\xi_{|E}$ has positive measure (and hence there are at most countably many of them).*

2. *$\xi_{|X\smallsetminus E}$ is isomorphic to the partition of the unit square with Lebesgue measure into vertical lines given in Example (2).*

*Proof.* The proof of this result is rather technical and we refer the interested reader to [CK] and [V]. □

[6] $\mu_\xi$ is often called *factor measure.*

## 3.5 Disintegration

If a partition element $C$ carries positive measure, then we can define a conditional measure on $C$ by the obvious method; given $E \subset C$, the conditional measure of $E$ is

$$\mu_C(E) := \frac{\mu(E)}{\mu(C)}. \tag{3.5.1}$$

However, for many partitions arising in the study of dynamical systems, we would also like to be able to define conditional measures on partition elements of zero measure, and to do so in a way which allows us to reconstruct the original measure. The model to keep in mind is the canonical example of a measurable partition, i.e. the square partitioned into vertical lines (like in example (2) of section 3.1). Then denoting by $\lambda$, $\lambda_1$ and $\lambda_2$ the Lebesgue measure on the square, the horizontal unit interval and vertical unit interval, respectively, Fubini's theorem says that for any integrable function $f : [0,1]^2 \to \mathbb{R}$ we have

$$\int_{[0,1]^2} f(x,y) d\lambda(x,y) = \int_{[0,1]} \int_{[0,1]} f(x,y) d\lambda_2(y) d\lambda_1(x). \tag{3.5.2}$$

By Theorem 3.4.5, any measurable partition of a Lebesgue space is isomorphic to the standard example; perhaps with a few elements of positive measure hanging about, but these will not cause any trouble, as we already know how to define conditional measure on them. Taking the pullback of the Lebesgue measures $\lambda_1$ and $\lambda_2$ (applied to a given vertical segment) under this isomorphism, we obtain a *factor measure* $\mu_\xi$ on $\xi$, which corresponds to the horizontal unit interval (the set of partition elements), and a *conditional measure* (on the pullback of that vertical segment) belonging to the family of conditional measures $\{\mu_C\}_{C\in\xi}$. To be precise, the previous statement is explicative but not exact: indeed, the set $\{\mu_C\}_{C\in\xi}$ of conditional measures is not defined *element-wise*. It does not make sense to speak of the conditional measure on a *single* element of the partition, because the conditional measures have a meaning only when integrated with respect to the factor measures. What we can say is that such an isomorphisms exists for almost every $C \in \xi$ with respect to the factor measure, and when we will speak about a system of conditional measures, we will refer to a set of measures defined on the element of a partition in this weaker sense. Note that the factor measure is exactly the measure on the space of partition elements which was described in the last section. Note also that although the measure $\lambda_2$ was the same for each vertical line (up to a horizontal translation), we can make no such statement about the measure $\mu_C$, as the geometry is lost in the purely measure theoretic isomorphism between $X$ and $[0,1]^2$. The key property of these measures is that for any integrable function $f : X \to \mathbb{R}$, the function

$$\xi \to \mathbb{R},$$

$$C \mapsto \int_C f d\mu_C$$

is measurable and we have

$$\int_X f d\mu = \int_\xi \int_C f d\mu_C d\mu_\xi. \tag{3.5.3}$$

Each $\mu_C$ is also "normalized" in the sense that $\mu_C(C) = 1$.
We cannot in general write a simple formula for the conditional measures, as we could in the case where partition elements carried positive weight, so on what grounds do we say that these conditional measures exist? The justification above relies on the characterization of measurable partitions given by Theorem 3.4.5. One can obtain the same result without constructing necessarily the isomorphism with the unit square.

**Definition 3.5.1.** Let $X$ be a compact metric space, $\mu$ a Borel probability measure on $X$, and $\xi$ be a partition of $X$ into measurable subsets. A *system of conditional measures of $\mu$ with respect to $\xi$* is a family $(\mu_C)_{C\in\xi}$ [7] of probability measures on $X$ such that

1. $\mu_C$ exists for $\mu_\xi$-almost every $C$ and $\mu_C(C) = 1$,
2. given any continuous function $\varphi : X \to \mathbb{R}$, the function $\xi \ni C \to \int \varphi d\mu_\xi$ is measurable and $\int \varphi d\mu = \int \big(\int \varphi d\mu_C\big) d\mu_\xi(C)$.

The result that states the existence of conditional measures is the following theorem by Rokhlin.

**Theorem 3.5.2.** *(Rokhlin) If $\xi$ is a measurable partition, then there exists s system of conditional measures of $\mu$ relative to $\xi$. Furthermore, such a decomposition is unique in the sense that if $\mu_C^1$ and $\mu_C^2$ are two conditional measure defined on the same element $C \in \xi$, then $\mu_C^1 = \mu_C^2$ for almost every $C$ with respect to the factor measure (note that for almost every $C \in \xi$ both $\mu_C^1$ and $\mu_C^2$ are defined).*

*Proof.* We refer the interested reader to the original paper of Rokhlin [R1] or the notes [V]. □

As a final remark we want to stress the following point. Looking at the defining formula for conditional measures (3.5.3), one may notice that conditional measures appear only integrated over the elements of the partition $\xi$. Indeed, we never deal with a conditional measure over a *single* partition element, except in the case of equation (3.5.1) of elements with positive measure. Nonetheless, one may wonder whether it is possible to define a conditional measure on a zero measure set (regardless of the problem of disintegration which is a different thing). To the eyes of an orthodox probabilist this might seem an ill-defined problem and he might, rightly, observe that null measure sets should play no role when considered individually. Anyway, if we want to do that there

[7] The conditional measure $\mu_C$ is defined only for $\mu_\xi$-almost every $C$, see also the discussion above about the meaning we give to a set of conditional measures.

is at least one canonical way based on *Hausdorff measure* and other parametrization dependent ways which are related to the so called *fan measures*, following the terminology of [BW]. There, the problem is addressed of conditioning on a null measure set and it is explained that there are different possible paradigms which lead to different results, that, fundamentally, can be all correct (even though there are some that are *more canonical* than others). In the same article, it is also explained that the fact that different parametrizations result in different conditional measure is intimately linked to the *Borel-Kolmogorov paradox*, which fascinates mathematicians since its discover by Kolmogorov 100 years ago.

# Chapter 4

# Billiards

In this chapter we want to present as directly as possible the theory of *LSUMs* (*local stable and unstable manifolds*). LSUMs are at the very foundation of the ergodic study of hyperbolic dynamical systems. In order to be as direct as possible in our exposition, we may lack of completeness in some part of it: the proofs of some theorems are only sketched, others are left to the references and some results are presented in an heuristic fashion. To remedy these shortcomings, the interested reader may look at the book of Chernov and Markarian [CM], which is a milestone in the field of mathematical billiards. Let us start by reminding the reader the definition of billiard dynamics already presented in the Introduction.

**Definition 4.0.1.** Let $\mathcal{D} \subset \mathbb{R}^2$ be a domain with smooth or piece-wise smooth boundary. A billiard system corresponds to a free motion of a point particle inside $\mathcal{D}$ with specular reflections off the boundary $\partial\mathcal{D}$.

The dynamics we have just described is represented by a flow in a Phase Space $\Omega$, that we will define later on. The flow simply represents the motion of the particle in the container and is generated by an Hamilonian with a singular potential (a rigid wall is in correspondence with an infinite barrier of potential). On the other hand, we will be mainly interested at the billiard map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ which is defined on a suitable Poincaré section $\mathcal{M}$ of $\Omega$, that we will define later as well. A typical $x \in \mathcal{M}$ posses two submanifolds (in the context of billiards, these are typically two smooth curves) that pass trough it: a local stable manifold (LSM) which comprises forward asymptotic points to $x$ and a local unstable manifold (LUM) made of backward asymptotic points to $x$. Focusing on LSM only (LUM are their time-reversal counterpart), we may say that they are the results of two processes: hyperbolicity and singularities. Roughly speaking, hyperbolicity tells us that for a typical point $x \in \mathcal{M}$ there exist two direction in the tangent space, namely $E^u_x$ and $E^s_x$ in which tangent vectors expand and contract, respectively, under the action of the differential of the billiard map $\mathcal{DF}$. For proving the existence of LSUMs this phenomenon is *good* because it tells us that there are certain directions in the phase space called *contracting* in which points get nearer and nearer one to the other as time elapses[1]. On the other side, singularities are *evil* because they separate the trajectories of points. If two points happen to be separated by a singularity, we can no longer say that they will get nearer one to each other as time elapses, even if the map is hyperbolic and they are in the right (contracting) direction. This endless war between hyperbolic contraction and singularities is won at time $n \to \infty$ by contraction. Indeed, we are able to find submanifolds $W^s$ of codimension one such that points get one near each other and are never detached by singularities. Nonetheless, the price to be paid for this victory is very high: there will be submanifolds $W^s$ with arbitrarily small diameter densely distributed in the phase space. Such a phenomenon creates countless problems in many proofs.

**Hopf Paths: showing ergodicity from the existence of LSUMs**

It is often said that hyperbolicity (which is linked to expansion in the some directions in the tangent space and sensitivity to initial conditions) produces chaos. In this section we briefly expose a method for proving ergodicity of a dynamical system endowed with an hyperbolic structure (i.e., existence of LSUMs at typical points).
Let $(M, d)$ be a metric space and $(M, \mathfrak{B}(M), m, T)$, $T : M \to M$ be a dynamical system on the measure space $(M, \mathfrak{B}(M), m)$, where $\mathfrak{B}(M)$ is the Borel $\sigma$-algebra on $(M, d)$ and $m$ is a probability measure, i.e., $m(M) = 1$. Assume that for $m$-almost every $x \in M$ there exists a LUM, call it $W^u(x)$ and a LSM, $W^s(x)$. For any $x$ for which it is defined, $W^s(x)$ has the property that for any $y, z \in W^s$ $d(T^n y, T^n z) \to 0$ as $n \to \infty$. Call an *observable* a measurable function $f : M \to \mathbb{R}$. Suppose that $f$ is also uniformly continuous[2]. Let us denote with $\bar{f}_+(x)$ and $\bar{f}_-(x)$ the forward and backward Birkhoff

[1] Note that hyperbolicity tells us only about contraction and expansion in the tangent space not on the Phase Space $\mathcal{M}$. On the other hand, LSUMs are realization of the same phenomena on $\mathcal{M}$. Heuristically, one can think of hyperbolicity as expansion or contraction at the first order of the Taylor expansion of $\mathcal{F}$, whereas LSUMs characterize expansion/contraction at all orders. Given that some map is hyperbolic, there exist classical theorems of hyperbolic dynamics which (under additional assumptions) prove the existence of LSUMs like the Hadamard-Perron theorem for the smooth case (see chapter 6 of [KH]) and Katok-Strelcyn theorems for smooth maps with singularities [KS]. See also the Hartman–Grobman theorem which links properties of a dynamical system in a domain near an equilibrium points to the linearized version of the same dynamical system, see [H]. We will try to give intuition of general methods in these proofs by addressing the more specific Sinai-billiard case later on.

[2] This request is not too restrictive because we can approximate e.g. $L^1$ functions with smooth compactly supported functions.

averages of $f$, respectively, i.e.,

$$\bar{f}_+(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} (f \circ T^n)(x),$$

$$\bar{f}_-(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} (f \circ T^{-n})(x),$$

The following proposition is a general fact in ergodic theory (see e.g. [CM], chapter 6).

**Proposition 4.0.2.** *Suppose $\bar{f}_+(x)$ is (mod 0) constant for every continuous function $f$. Then the map $T$ is ergodic. (The converse also holds.)*

By the uniform continuity of $f$ one has that $|f(T^n(x)) - f(T^n(y))| \to 0$ for any $x, y \in W^s$, as $n \to \infty$. This means that $\bar{f}_+$ is constant in $W^s$. By the time reverse argument one has that $\bar{f}_-$ is constant on any LUM $W^u$. Furthermore, one can prove that $\bar{f}_+(x) = \bar{f}_-(x)$ for almost every $x \in M$. Now, the general idea for proving ergodicity for the dynamical system $(M, \mathfrak{B}(M), m, T)$ is proving the existence of a *Hopf Path* that links $x$ and $y$ for almost every pair of points $x, y \in M$.

*By a Hopf Path linking $x$ and $y$ we mean an alternating sequence of LSMs and LUMs $\{W_i\}_{i=1}^N$ such that $x \in W^1$, $y \in W^N$ and $W^i \cap W^{i+1} = \{x_i\}$ where $x_i$ are typical points in the sense that $\bar{f}_+(x_i) = \bar{f}_-(x_i)$. (See the picture below)*

$M$

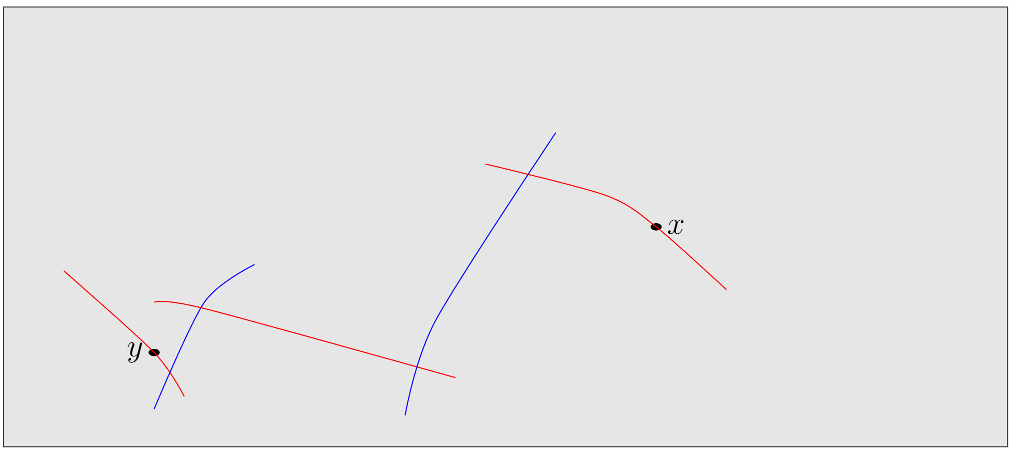


Figure 3: an Hopf Path connecting $x$ with $y$. LUMs are in blue while LSMs are in red.

If we are able to connect two typical points $x$ and $y$ with an Hopf path, we know that they belong the same ergodic components (this is a consequence of Proposition 4.0.2). Therefore, we prove ergodicity of the whole dynamical system by showing the existence of an Hopf Path between any two pairs of points taken from a full measure set. For real smooth dynamical systems with singularities (like billiards), a great difficulty is precisely to show that LSUMs are in a certain sense big enough to have Hopf Paths for almost every couple of points. The most technical and hard part of the proofs on the existence of Hopf Paths are the so called *Local Ergodic Theorems* (LETs), which establish the existence of *good points* that posses a neighborhood belonging to the same ergodic component[3]. For modern LETs, we refer the interested reader to [LW] and [DM]. Later on, we will see that properties of LUMs (and of their future and past dynamics) enable us to prove even more strong chaotic properties of the map, like the Bernoulli property (only for the finite measure case) and K-mixing.

**Billiards and smooth maps with singularities that preserves an infinite measure: a quick snapshot toward the original part of this thesis.**

The Structure Theorem we formulate in Chapter 6 holds for recurrent smooth maps with singularities which preserves a, possibly infinite, measure (for all the assumptions we refer to Chapter 6). It is a fact that this kind of results find most of their applications in Billiard dynamics[4]. However, we want to stress that in that context applications of our theorem are quite broad since it applies to Lorentz Gases and Tubes even in a quenched dynamical setup and in higher dimensions (see the last section of Chapter 6 for a more complete discussions on the applications and for the relevant references). Furthermore, for the aforementioned dynamical systems (which preserve an infinite measure), the conclusions we formulate are, to our knowledge, new. On the other hand, in the present chapter (which should be considered as introductory material), we present the theory of LSUMs for planar, dispersing, finite-measure Sinai billiards. We do so because here we just want to give a feel to the reader of what LSUMs are and what typical properties they enjoy. Therefore, to be as clear and direct as possible we present the simplest (nonetheless representative) case. At any rate, we stress that Chapter 6 is self-sufficient and all the relevant definitions, assumptions and preliminary results are contained there or in the works mentioned therein.

[3] Then typically one uses density of good points and global properties of the map to prove ergodicity.

[4] We will see that the billiard map $\mathcal{F}$ preserves a measure induced by the Liouville measure preserved by the Billiard flow.

## 4.1 Billiard tables

In this section, we give the correct definitions for billiard dynamics in the simpler case of Sinai billiards. As we observed in the previous section, we stress that the billiard-like systems that we study in the original part of this thesis (in Chapter 6) are more complicated than the simple models presented here. We do so to give a quick introduction to the subject. Although Definition 4.0.1 is very general and leaves admissible an incredible variety of different types of dynamical behavior, we restrict ourselves to billiards which are in some sense less pathological than others (see assumptions A1-A4). Among this class of *admissible* billiards we will mainly consider the simplest possible ones, which looks pretty much like this:

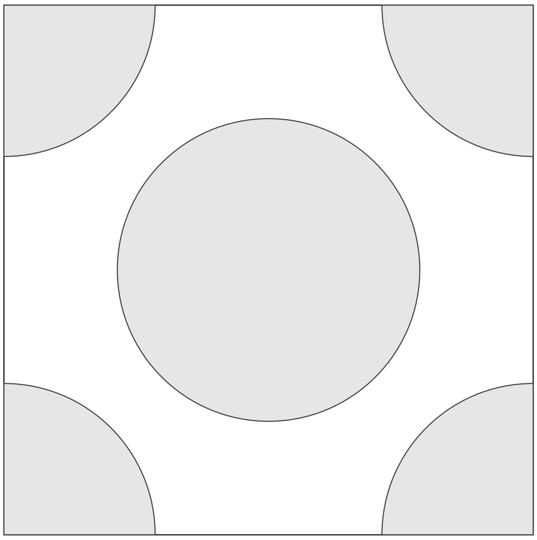

Figure 4: an example of the 'simple' model of dispersing billiard on the torus.

Let $\mathcal{D}_0 \subset \mathbb{R}^2$ or $\mathcal{D}_0 \subset \mathbb{T}^2$ be a bounded open connected domain and $\mathcal{D} = \bar{\mathcal{D}}_0$ denote its closure.

ASSUMPTION A1. The boundary $\partial\mathcal{D}$ is a finite union of smooth ($C^l$, $l \geq 3$) compact curves:

$$\partial D := \Gamma := \Gamma_1 \cup ... \cup \Gamma_r. \tag{4.1.1}$$

Precisely, each curve $\Gamma_i$ is defined by a $C^l$ map $f_i : [a_i, b_i] \to \mathbb{R}^2$, which is one-to-one on $[a_i, b_i)$ and has one sided derivatives, up to order $l$, at points $a_i$ and $b_i$. The value $l$ is the class of smoothness of the billiard table. The curves $\Gamma_i$ are called *walls of the table* $\mathcal{D}$.

ASSUMPTION A2. The boundary components $\Gamma_i$ can intersect each other only at their endpoints;

$$\Gamma_i \cap \Gamma_j \subseteq \partial\Gamma_i \cup \partial\Gamma_j \quad \text{for } i \neq j. \tag{4.1.2}$$

We put

$$\Gamma_* = \partial\Gamma_1 \cup ... \cup \partial\Gamma_r \quad \text{ and } \quad \tilde{\Gamma} = \Gamma \setminus \Gamma_*. \tag{4.1.3}$$

We call $x \in \Gamma_*$ *corner points* of $\mathcal{D}$ and $x \in \tilde{\Gamma}$ *regular boundary points.*

We fix an orientation of each $\Gamma_i$ so that $\mathcal{D}$ *lies to the left* of $\Gamma_i$. Then we parametrize every $\Gamma_i$ by its arclenght; thus tangent vectors become unit vectors: $||f_i'|| = 1$.

ASSUMPTION A3. On every $\Gamma_i$, the second derivative $f_i''$ either never vanishes or is identically zero (thus, every wall $\Gamma_i$ is either a curve without inflection points or a line segment).

Note that $f_i'' \perp f_i'$ because $||f_i'|| = const$. Hence, if $f_i'' \neq 0$, then the pairs of vectors $f_i'$ and $f_i''$ is either left or right along the curve $\Gamma_i$. Accordingly, we can distinguish three types of walls:
**Flat walls**: such that $f'' \equiv 0$;
**Focusing walls:** such that $f'' \neq 0$ is pointing inside $\mathcal{D}$;
**Dispersing walls:** such that $f'' \neq 0$ is pointing outside $\mathcal{D}$.
Following [CM] we define the *(singed) curvature* on each $\Gamma_i$ as follows:

$$\mathcal{K} = \begin{cases} 0 & \text{if } \Gamma_i \text{ is flat,} \\ -||f''|| & \text{if } \Gamma_i \text{ is focusing,} \\ ||f''|| & \text{if } \Gamma_i \text{ is dispersing.} \end{cases} \tag{4.1.4}$$

Accordingly, we define

$$\Gamma_0 = \bigcup_{\mathcal{K}=0} \Gamma_i, \quad \Gamma_- = \bigcup_{\mathcal{K}<0} \Gamma_i \quad \Gamma_+ = \bigcup_{\mathcal{K}>0} \Gamma_i. \tag{4.1.5}$$

We note that the curvature of each focusing and dispersing wall $\Gamma_i$ is bounded away from zero and infinity, due to assumption A3 and the closedness of $\Gamma_i$. We denote it by $|\Gamma_i|$ and set $|\Gamma| = \sum_i |\Gamma_i|$ be the total perimeter of $\mathcal{D}$. A clever way to generate finite measure billiards is to put a scatter (some closed two-dimensional figure) on the two-dimensional torus $\mathbb{T}^2$ (see e.g. figure 4). We want to include this possibility, too. We summarize our constructions as follows:

**Definition 4.1.1.** A *billiard table* $\mathcal{D}$ is the closure of a bounded open connected domain $\mathcal{D} \subset \mathbb{R}^2$ or $\mathcal{D} \subsetneq \mathbb{T}^2$ such that $\partial\mathcal{D}$ satisfy assumption A1-A3.

## 4.2 Billiard map

Here we construct the dynamics of a billiard table trough the billiard flow. This is not going to be a simple task, since there will be several cases where our construction fails and the trajectory of a billiard particle cannot be defined (this is the way in which singularities arise). Let $q \in \mathcal{D}$ denote the position of the moving particle and $v \in \mathbb{R}^2$ its velocity vector. Clearly both $q$ and $v$ are function of time $t \in \mathbb{R}$, i.e. $q = q(t)$ and $v = v(t)$. We have two equations that specify the dynamics. When $q \in \mathcal{D}$, the particle is inside the table and

$$\dot{q} = v \quad and \quad \dot{v} = 0 \tag{4.2.1}$$

(here the dot means time derivative). When $q \in \tilde{\Gamma}$, the particles experience a regular collision and

$$v^+ = v^- - 2 < v, n > n, \tag{4.2.2}$$

where $v^+$ and $v^-$ refer to the postcollisional and precollisional velocities, respectively, and $n$ denotes the unit normal vector to $\tilde{\Gamma}$ at the point $q$. Equation (4.2.2) is nothing more than the classical rule "the angle of incidence is equal to the angle of reflection". If the moving particle hits a corner point, i.e. $q \in \Gamma_*$, it stops, and its motion will no longer be defined beyond that point. The equations (4.2.1) and (4.2.2) preserve the norm $||v||$, and its customary to set it to one: $||v|| = 1$. For reason that will be clear when we will introduce the *billiard map*, one wants to exclude *grazing collisions* from the *regular collisions.*

**Definition 4.2.1.** A collision is said to be *regular* if $q \in \tilde{\Gamma}$ and the vector $v^-$ is not tangent to $\Gamma$. In this case $v^+ \neq v^-$. If $v^-$ is tangent to $\Gamma$ at the point of collision, then $v^+ = v^-$, and such a collision is said *grazing* or *tangential.*

Clearly, grazing collisions can occur only at dispersing walls. Simple geometrical considerations imply also the following fact.

**Proposition 4.2.2.** *Let the moving particle collide with the regular part of the boundary at time $t$ (that is $q(t) \in \tilde{\Gamma}$). Then it will move inside $\mathcal{D}$ without collisions during some time interval $(t, t+\epsilon)$ for some positive $\epsilon$.*

We now summarize our observations:

**Proposition 4.2.3.** *The trajectory of the particle $(q(t), v(t))$ starting at $q(0) \in int\mathcal{D}$ is defined at all times $-\infty < t < \infty$ unless one of the two exceptions occurs:*

*a) The particle hits a corner point of $\mathcal{D}$; i.e. $q(t) \in \Gamma_*$ for some $t \in \mathbb{R}$.*

*b) Collision times $\{t_n\}$ have an accumulation point in $\mathbb{R}$.*

The second exception mentioned in Proposition 4.2.3 is quite particular and we want to exclude it from our study. It turns out that under our assumptions (A1)-(A3) it may be the case[5] that condition b) takes place only if we have a cusp made by a focusing wall and a dispersing wall. Therefore, we introduce the following assumption.

ASSUMPTION A4. Any billiard table $\mathcal{D}$ contains no cusps made by a focusing and a dispersing walls.

### 4.2.1 Phase space for the flow

The *state* of the moving particle at any time is specified by its position $q \in \mathcal{D}$ and the unit vector $v \in S^1$. Thus, the *phase space* of the system is

$$\Omega = \{(q, v)\} = \mathcal{D} \times S^1.$$

This is a three dimensional manifold with boundary $\partial\Omega = \Gamma \times S^1$. One can picture $\Omega$ as a "doughnut" whose crosse-section is $\mathcal{D}$. Furthermore, at each regular boundary point $q \in \tilde{\Gamma}$ it is convenient to identify the pairs $(q, v^-)$ and $(q, v^+)$ related by the collision rule (4.2.2), which amounts to "gluing" $\Omega$ along its boudary. Let $\tilde{\Omega} \subseteq \Omega$ denotes the set of states $(q, v)$ on which the dynamics of the moving particle is defined at all times $-\infty < t < \infty$. We obtain a one-parameter group of transformation (flow)

$$\Phi^t : \tilde{\Omega} \to \tilde{\Omega},$$

with continuous $t \in \mathbb{R}$. That is, $\Phi^0 = Id$ and $\Phi^{t+s} = \Phi^t \circ \Phi^s$ for all $t, s \in \mathbb{R}$. Let $\tilde{\Omega} = \tilde{\Omega}_c \cup \tilde{\Omega}_f$, where $\tilde{\Omega}_c$ contains all trajectories with collisions and $\tilde{\Omega}_f$ is the union of all collision free trajectory. Clearly, $\tilde{\Omega}_c$ and $\tilde{\Omega}_f$ are both invariant under the flow. A simple and useful fact is the following Proposition.

**Proposition 4.2.4.** *If $\mathcal{D}$ is a billiard table in $\mathbb{T}^2$, then every trajectory of the flow experiences either infinitely many collisions or none at all.*

*Proof.* It is a well-known fact that every free trajectory on $\mathbb{T}^2$ is either periodic or dense (see e.g. [KH], chapter 1). Then, supposing that a moving particle has experienced some collision in the past and will not reflect anymore at the boundary, it will proceed along its free motion, passing exactly at or arbitrary near the place of its last collision without bouncing off the walls anymore, which is absurd. $\square$

### 4.2.2 Collision map

Here we describe the object that interests us most: the *collision map* or *billiard map*. There is a standard way to reduce a flow to a map by constructing a cross-section. Given

[5] If it *really* is the case is still an open question, see [CM], section 2.4.

a flow $\Phi^t$ on a manifold $\Omega$, one finds a hypersurface $M \subset \Omega$ transversal to the flow so that each trajectory crosses $M$ infinitely many times. Then the flow induces a return map $F : M \to M$ and a return time function $L(x) = \min\{s > 0 : \Phi^s(x) \in M\}$ on $M$, so that $F(x) = \Phi^{L(x)}(x)$.
For a billiard system, the hypersurface in $\Omega$ is usually taken to be the set $\Gamma \times S^1$, i.e. the boundary of the table with the velocity. There is a delicate part, however, in this construction, since we have identified the pre- and post-collisional velocity vectors $v^-$ and $v^+$, transforming effectively $S^1$ into an half-circle. To take this into account, it is customary to describe the cross-section as the set of all possible postcollisional velocity vectors:

$$\mathcal{M} = \bigcup_i \mathcal{M}_i, \quad \mathcal{M}_i = \{x = (q, v) \in \Omega : q \in \Gamma_i, < v, n > \geq 0\}, \tag{4.2.3}$$

where $n$ denotes the normal vector to $\Gamma_i$ pointing inside $\mathcal{D}$. The set $\mathcal{M}$ is a two-dimensional submanifold in $\Omega$ called *collision space*. If the trajectory $\Phi^t x$ for $x \in \mathcal{M}$ is defined during some interval $(0, \epsilon)$ then, by Proposition 4.2.4, it must intersect the surface $\mathcal{M}$ at a future time $\tau(x)$, called *return time*. Any trajectory of the flow $\Phi^t : \tilde{\Omega}_c \to \tilde{\Omega}_c$ crosses the surface $\mathcal{M}$ infinitely many times. Let

$$\tilde{\mathcal{M}} = \mathcal{M} \cap \tilde{\Omega} = \mathcal{M} \cap \tilde{\Omega}_c. \tag{4.2.4}$$

This defines the *return map* or *billiard map*

$$\mathcal{F} : \tilde{\mathcal{M}} \to \tilde{\mathcal{M}} \quad by \quad \mathcal{F}(x) = \Phi^{\tau(x)+0} x. \tag{4.2.5}$$

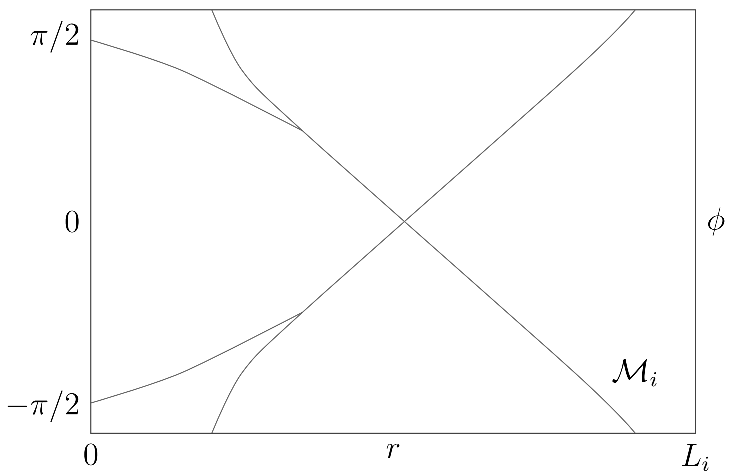


Figure 5: a subset $\mathcal{M}_i$ of the boundary $\mathcal{M}$. In the figure, we have plot also lines representing the singularity set $\mathcal{S}_1 \cup \mathcal{S}_{-1}$ that we will define in the next section.

### 4.2.3 Coordinates for the map and its singularities

Let $r$ be a fixed arc length parameter on each $\Gamma_i$ so that $r$ takes value in an interval $[a_i, b_i]$. For each point $x \in \mathcal{M}$, let $\phi \in [-\pi/2, \pi/2]$ denote the angle between $v$ and $n$ oriented clockwise.. Then $r$ and $\phi$ make coordinates on $\mathcal{M}$. For each smooth curve $\Gamma_i$ the manifold $\mathcal{M}_i = \Gamma_i \times [-\pi/2, \pi/2]$ is a rectangle. We denote

$$\mathcal{S}_0 := \partial\mathcal{M} = \{|\phi| = \pi/2\} \cup \Big(\cup_i(\{r = a_i\} \cup \{r = b_i\})\Big), \tag{4.2.6}$$

where the set $\{r = a_i\} \cup \{r = b_i\}$ is included only for $\Gamma_i$'s which are *not* smooth closed curves (i.e. when it is a *true* boundary of the interval $[a_i, b_i]$). For every point $x \in int\mathcal{M}$ its trajectory $\Phi^t(x)$ is defined, at least, for $0 < t < \tau(x)$, i.e. until the next intersection with $\mathcal{M}$, at which we have three possible cases:

a) a regular collision, i.e. $\mathcal{F}(x) \in int\mathcal{M}$;

b) a grazing collision at a dispersing wall, i.e. $\mathcal{F}(x) \in \mathcal{S}_0$;

c) the trajectory hits a corner and dies.

In the last case $\mathcal{F}(x)$ is not defined. We define

$$\mathcal{S}_1 := \mathcal{S}_0 \cup \{x \in int\mathcal{M} : \mathcal{F}(x) \notin int\mathcal{M}\}, \tag{4.2.7}$$

where the second set in the union coincides with points where event b) or c) occurs. Analogously

$$\mathcal{S}_{-1} := \mathcal{S}_0 \cup \{x \in int\mathcal{M} : \mathcal{F}^{-1}(x) \notin \mathcal{M}\}. \tag{4.2.8}$$

We regard $\mathcal{S}_1$ as the singularity set of the map $\mathcal{F}$ and $\mathcal{S}_{-1}$ as the singularity set for $\mathcal{F}^{-1}$. 'Morally' one can think that $\mathcal{S}_0 = \partial\mathcal{M}$, $\mathcal{S}_1 = \mathcal{S}_0 \cup$ '$\mathcal{F}^{-1}(\partial\mathcal{M})$' and $\mathcal{S}_{-1} = \mathcal{S}_0 \cup$ '$\mathcal{F}(\partial\mathcal{M})$', even if the last two expressions are ill-defined. Excluding those singularity points the map $\mathcal{F}$ is smooth in the following sense.

**Theorem 4.2.5.** *The map $\mathcal{F} : \mathcal{M} \setminus \mathcal{S}_1 \to \mathcal{M} \setminus \mathcal{S}_{-1}$ is a $C^{l-1}$ diffeomorphism. Furthermore, let $\mathcal{F}(r, \phi) = (r_1, \phi_1)$ for any $(r, \phi) \in \mathcal{M}$. Then, the differential $\mathcal{DF}$ of the map $\mathcal{F}$ at the point $x = (r, \phi) \in \mathcal{M}$ is expressed in the $(r, \phi)$ coordinates by the following $2 \times 2$ matrix*

$$\mathcal{D}_x\mathcal{F} = \frac{-1}{\cos\phi_1}\begin{bmatrix} \tau\mathcal{K} + \cos\phi & \tau \\ \tau\mathcal{K}\mathcal{K}_1 + \mathcal{K}\cos\phi_1 + \mathcal{K}_1\cos\phi & \tau\mathcal{K}_1 + \cos\phi_1, \end{bmatrix} \tag{4.2.9}$$

*where $\tau = \tau(x)$, $\mathcal{K} = \mathcal{K}(x)$ and $\mathcal{K}_1 = \mathcal{K}(\mathcal{F}(x))$.*

The expression (4.2.9) shows that the derivatives of $\mathcal{F}$ are unbounded: they blow up as $cos\phi_1 \to 0$, i.e. when $x_1$ is near $\mathcal{S}_0$ (or, equivalently, $x$ is near $\mathcal{S}_1$). A deeper study of singularities will be taken over in a later section.

### 4.2.4 Invariant measure of the map

Among all the dynamical system, those who preserves a measure plays a very special role. Using (4.2.9), one can see that billiard system do preserve a mesure.

**Proposition 4.2.6.** *The map $\mathcal{F}$ preserves the measure $cos\phi dr d\phi$ on $\mathcal{M}$.*

*Proof.* By (4.2.9),

$$det\mathcal{D}_x\mathcal{F} = \frac{\cos\phi}{\cos\phi_1}.$$

Using this and changing variable give

$$\iint_{\mathcal{F}(A)} \cos\phi_1 dr_1 d\phi_1 = \iint_A cos\phi dr d\phi,$$

for any Borel set $A \subseteq \mathcal{M}$. □

Note that

$$\iint_{\mathcal{M}} cos\phi dr d\phi = \int_{-\pi/2}^{\pi/2} cos\phi d\phi \int_\Gamma dr = 2|\Gamma|.$$

We define the *canonical probability measure* $\mu$ on $\mathcal{M}$ as the one which is preserved by $\mathcal{F}$ and is normalized:

$$d\mu = (2|\Gamma|)^{-1} cos\phi dr d\phi. \tag{4.2.10}$$

We observe that $\mu$ is equivalent to the Lebesgue measure $dLeb = drd\phi$ on $\mathcal{M}$. Indeed, it has with respect to it the almost everywhere positive density $cos\phi$.

### 4.2.5 Involution

The dynamics of billiards is special because it is inspired by Physics. Because of that it has some additional symmetries that may not be owned by general measure-preserving invertible transformation. For example, they have the important property of *time reversibility* or *involution property.* For any $x = (q, v) \in \Omega$ the point $\mathcal{I}_\Omega(x) = (q, -v)$ satisfies

$$\Phi^{-t}(\mathcal{I}_\Omega(x)) = \mathcal{I}_\Omega(\Phi^t x)$$

whenever the flow is defined. We can loosely say that the involution map $\mathcal{I}_\Omega : \Omega \to \Omega$ commutes with the flow, i.e.

$$\Phi^{-t} \circ \mathcal{I}_\Omega = \mathcal{I}_\Omega \circ \Phi^t.$$

This means, plainly, that if we reverse the particle's velocity, it will retrace its past trajectory 'backwards'. Since the collision map is just a cross section of the flow, it will also enjoy the involution property, i.e.

$$\mathcal{F}^{-k} \circ \mathcal{I} = \mathcal{I} \circ \mathcal{F}^k, \quad k \in \mathbb{Z},$$

(whenever the trajectory until time $k$ of $x$ is defined) where $\mathcal{I} : \mathcal{M} \to \mathcal{M}$ is defined by $\mathcal{I}(x) = (r, -\phi)$ for any $x = (r, \phi) \in \mathcal{M}$. Note that $\mathcal{I}$ preserves the measure $\mu$. Here there is a little consequence of our definitions:

**Proposition 4.2.7.** *Consider the vector function $\vec{L}(r, \phi) = (\tau \cos \omega, \tau \sin \omega)$, where $\tau = \tau(x)$ is the return time and $\omega$ is the angle between the outgoing velocity vector immediately after the collision and the $x$ axis. $\vec{L}$ is the first directed link of the billiard trajectory originating from $(r, \phi)$. Then*

$$\int_{\mathcal{M}} \vec{L} d\mu = 0.$$

*In other words, doing the mean of all the possible links between points of the boundary $\partial\mathcal{D}$ we have zero.*

*Proof.* Note that $\vec{L}(x) = -\vec{L}(\mathcal{I} \circ \mathcal{F}(x))$. Using the invariance of $\mu$ both under $\mathcal{F}$ and $\mathcal{I}$

$$\int_{\mathcal{M}} \vec{L} d\mu = \int_{\mathcal{M}} \vec{L}(\mathcal{I} \circ \mathcal{F}(x)) d\mu = -\int_{\mathcal{M}} \vec{L} d\mu.$$

Hence

$$\int_{\mathcal{M}} \vec{L} d\mu = 0.$$

Indeed, for any trajectory we have also the backward counterpart is admissible and, doing the vector sum, the final result is zero. □

What is different from zero is the mean return time

$$\bar{\tau} := \int_{\mathcal{M}} \tau(x) d\mu(x).$$

There exists a beautiful formula (known also in the filed of integral geometry) that characterize $\bar{\tau}$ using just the area $|\mathcal{D}|$ of the table and its perimeter $|\Gamma|$. It says that, simply,

$$\bar{\tau} = \frac{\pi |\mathcal{D}|}{|\Gamma|}.$$

One can also prove this using the Birkhoff theorem, see section 2.12 of [CM].

## 4.3 Lyapunov exponents and hyperbolicity

Lyapunov exponents play a major role in the study of differentiable conservative dynamical system; they measure the stability (or instability) of trajectories under small perturbation. The mathematical way to codify the term *small perturbation* is thinking of them as vectors in the tangent space. This explains why Lyapunov exponents are

defined only when the dynamics takes place on a manifold. Suppose we follow a point during its trajectory. If there was associated to the initial point an error and the dynamics is deterministic, it is clear that also all the future images of that point will be associated to some error. How these errors are related in magnitude? Do they grow? At which rate? A greater than zero Lyapunov exponent means precisely that the error of an initial data 'propagates' exponentially fast in the asymptotic future.
Let $M$ be a compact Riemannian manifold (pheraps, with boundary and corners), $N \subseteq M$ an open and dense subset and $F : N \to M$ a $C^r$ (with $r \geq 2$) diffeomorphism of $N$ onto $F(N)$. Note that all the iterations of $F$ are defined on the set

$$\tilde{N} = \bigcap_{n=-\infty}^{\infty} F^n(N).$$

Assume that $F$ preserves a probability measure $\mu$ on $M$ and $\mu(\tilde{N}) = 1$. Let $||-||$ be the Euclidean norm on $\mathcal{M}$.

**Theorem 4.3.1.** *(Oseledets). Suppose*

$$\int_M \log^+ ||D_x F|| d\mu(x) < \infty \quad \textit{and} \quad \int_M \log^+ ||D_x F^{-1}|| d\mu(x) < \infty,$$

*where* $\log^+(s) = \max\{\log s, 0\}$*. Then there exists an* $F$*-invariant set* $H \subset \tilde{N}$, $\mu(H) = 1$, *such that for all* $x \in H$ *there exists a* $DF$*-invariant decomposition of the tangent space*

$$\mathcal{T}_x M = E_x^{(1)} \oplus ... \oplus E_x^{(m)}, \tag{4.3.1}$$

*with some* $m = m(x)$*, such that for all nonzero vectors* $v \in E_x^{(i)}$

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{||D_x F^n v||}{||v||} = \lambda_x^{(i)}, \tag{4.3.2}$$

*where* $\lambda_x^{(1)} > ... > \lambda_x^{(m)}$.

The values $\lambda_x^{(i)}$ are called *Lyapunov exponents* of the map $F$ at the point $x$ and $k_i = \dim E_x^i$ their *multiplicities*. It immediately follows from (4.3.2) that the Lyapunov exponents and their multiplicities are invariant under the map $F$. Therefore, if the map $F$ is ergodic, the Lyapunov exponents $\lambda^{(1)} > ... > \lambda^{(m)}$ as well as their multiplicities are almost everywhere constant.

**Definition 4.3.2.** A point $x \in M$ is said to be *hyperbolic* if Lyapunov exponents exist at $x$ and none of them equals zero. For a hyperbolic point $x \in M$, we set $\mathcal{T}_x M = E_x^u \oplus E_x^s$, where

$$E_x^u = \bigoplus_{\lambda_x^{(i)}>0} E_x^{(i)} \quad \textit{and} \quad E_x^s = \bigoplus_{\lambda_x^{(i)}<0} E_x^{(i)} \tag{4.3.3}$$

(the superscript $u$ and $s$ stand for *unstable* and *stable*, respectively).

Equation (4.3.2) can be loosely restated as

$$||D_x F^n v|| \sim e^{n\lambda_x^{(i)}} ||v||.$$

Even if we will give a more precise characterization later, this is enough to understand the generic behavior of vectors in the subspaces $E_x^u$ and $E_x^s$. The first ones increase exponentially and the second ones decrease exponentially, in the asymptotic future. Note that a vector in the unstable subspace can contract for an arbitrary large amount of time before entering in the expanding regime and, viceversa, a stable vector can expand for an arbitrary large time in the future! Lyapunov exponents tell us only what happens in the *limit* $t \to \pm\infty$, but everything could happen 'before the limit'.

**Definition 4.3.3.** A map $F$ is said to be *hyperbolic* if $\mu$-almost every point $x \in M$ is hyperbolic.

Oseledets theorem do not say whether a map is hyperbolic, it just say that Lyapunov exponents are well defined but some of them (or all of them, eventually) can be zero. Hyperbolicity is a dynamical property but we will see that it has deep consequences in the ergodic and stochastic properties of a large class of maps. Suppose, instead, that we know that some given map is hyperbolic. Then we know that for almost every point we can look at a stable direction and at an unstable one. Nevertheless, we don't know how this *field of directions* is distributed in the phase space, so that it is possible that the stable direction approaches the unstable one in some area of the phase space. This situation can be dangerous because we could arrive to a point in which we cannot *effectively* distinguish between the stable and unstable directions. Luckily, we have a theorem by [KS] that guarantees that such an approach can happen only at sub-exponential speed, which is a useful addition to Oseledets theorem.
Pick $a \in \mathbb{R}$ such that Lyapunov exponents are different from $a$ at $\mu$-almost every point $x \in M$ Denote

$$E_x^{a+} = \bigoplus_{\lambda_x^{(i)} > a} E_x^{(i)} \quad \textit{and} \quad E_x^{a-} = \bigoplus_{\lambda_x^{(i)} < a} E_x^{(i)}, \tag{4.3.4}$$

and let $\gamma_a(x)$ denote the angle between the spaces $E_x^{a+}$ and $E_x^{a-}$.

**Proposition 4.3.4.** *For $\mu$-almost every point $x \in M$ we have*

$$\lim_{n \to \pm\infty} \frac{1}{n} \gamma_a(F^n(x)) = 0. \tag{4.3.5}$$

In other words, the angle $\gamma_a(F^n(x))$ may approach zero, as $n \to \infty$, but it does so more slowly than any exponential function. Lastly, we make an important distinction between *uniform* and *nonuniform* hyperbolicity. The following proposition will help us to approach this issue. Suppose $x \in M$ is a hyperbolic point and set

$$\lambda_x = \min_i |\lambda_x^{(i)}| > 0.$$

**Proposition 4.3.5.** *For any small $\epsilon > 0$ there is a $C(x,\epsilon) > 0$ such that for all $n \geq 1$*

$$||D_xF^{-n}(v)|| \leq C(x,\epsilon)e^{-n(\lambda_x-\epsilon)}||v|| \quad \forall v \in E^u_x \tag{4.3.6}$$

*and*

$$||D_xF^{n}(v)|| \leq C(x,\epsilon)e^{-n(\lambda_x-\epsilon)}||v|| \quad \forall v \in E^s_x. \tag{4.3.7}$$

*Proof.* First we show that for any given nonzero vector $v \in E^u_x$ (respectively, $v \in E^s_x$), there is a $C(x,\epsilon)$ satisfying (4.3.6) and (4.3.7) but it might depend on $v$, i.e. $C(x,\epsilon) = C(x,\epsilon,v)$. By (4.3.2), we have that for any $\epsilon > 0$ and any $v \in E^u_x$ there exists an $N \in \mathbb{N}^+$ such that

$$\frac{1}{n}\log\frac{||D_xF^nv||}{||v||} \geq (\lambda_x - \epsilon) \quad \forall n \geq N. \tag{4.3.8}$$

Therefore, we have

$$\frac{||D_xF^nv||}{||v||} \geq e^{n(\lambda_x-\epsilon)} \quad \forall n \geq N. \tag{4.3.9}$$

Since in Equation (4.3.9) both members are positive, there exists a constant, which we call $C(x,\epsilon,v)^{-1}$, such that

$$\frac{||D_xF^nv||}{||v||} \geq C(x,\epsilon,v)^{-1}e^{n(\lambda_x-\epsilon)} \quad \forall n \geq 1. \tag{4.3.10}$$

Evaluating (4.3.10) for the vector $D_xF^{-n}v \in E^u_{F^{-n}(x)}$ gives

$$||D_xF^{-n}v|| \leq C(x,\epsilon,v)e^{-n(\lambda_x-\epsilon)}||v|| \quad \forall n \geq 1. \tag{4.3.11}$$

Let us denote by $k$ the dimension of the unstable subspace $E^u_x$, i.e. $k = \dim E^u_x$. If we pick an orthonormal basis $e_1, ..., e_k$ of $E^u_x$ let us define

$$\tilde{C}(x,\epsilon) := \max_i\{C(x,\epsilon,e_i)\},$$

so that, for any $1 \leq i \leq k$,

$$||D_xF^{-n}e_i|| \leq C(x,\epsilon)e^{-n(\lambda_x-\epsilon)} \quad \forall n \geq 1. \tag{4.3.12}$$

Set

$$C(x,\epsilon) := k\tilde{C}(x,\epsilon).$$

Now, let us expand any $v \in E^u_x$ as $v = \sum_{i=1}^k a_iv_i$ where $a_i \in \mathbb{R}$. Then, using triangular inequality

$$\begin{aligned}||D_xF^{-n}v|| &\leq \Big(\sum_{i=1}^k |a_i|\Big)||D_xF^{-n}e_i|| \leq \Big(\sum_{i=1}^k |a_i|\Big)\tilde{C}(x,\epsilon)e^{-n(\lambda_x-\epsilon)} \leq k\tilde{C}(x,\epsilon)e^{-n(\lambda_x-\epsilon)}||v|| = \\ &= C(x,\epsilon)e^{-n(\lambda_x-\epsilon)}||v||.\end{aligned} \tag{4.3.13}$$

Equation (4.3.7) is deduced in a similar fashion. □

If $F$ is an hyperbolic map, then $\lambda_x > 0$ and $C(x, \epsilon) > 0$ are measurable functions on $M$. We say that the hyperbolicity is *uniform* if these functions can be made constant.

**Definition 4.3.6.** A hyperbolic map $F$ is *uniformly hyperbolic* if there are $\lambda$ and $C > 0$ such that for every $x \in M$ and all $n \geq 1$

$$||D_x F^{-n}(v)|| \leq Ce^{-n\lambda}||v|| \quad \forall v \in E_x^u \tag{4.3.14}$$

and

$$||D_x F^{n}(v)|| \leq Ce^{-n\lambda}||v|| \quad \forall v \in E_x^s, \tag{4.3.15}$$

whenever $E_x^u$ and $E_x^s$ exist.

### 4.3.1 Lyapunov exponents for the map

Here we apply the general theory of Lyapunov exponents to the billiard map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ preserving the measure $\mu$.

**Theorem 4.3.7.** *The Oseledets theorem applies to the billiard map $\mathcal{F}$, so Lyapunov exponents exist at $\mu$-almost every point $x \in \mathcal{M}$.*

*Proof.* We need to verify that the functions $\log^+ ||\mathcal{D}_x\mathcal{F}||$ and $\log^+ ||\mathcal{D}_x\mathcal{F}^{-1}||$ are integrable over $\mathcal{M}$. By the involution property, it is enough to do this only for $\log^+ ||\mathcal{D}_x\mathcal{F}||$. It follows from (4.2.9) that

$$\log^+ ||\mathcal{D}_x\mathcal{F}|| \leq \frac{C}{cos\phi_1},$$

where $C > 0$ is a constant (note that the curvature $\mathcal{K}$ of the wall $\Gamma$ is bounded). Then

$$\int_{\mathcal{M}} \log^+ ||\mathcal{D}_x\mathcal{F}|| \leq \int_{\mathcal{M}} |\log C + \log\cos\phi_1| d\mu \leq |\log C| + (2|\Gamma|)^{-1} \int_{\mathcal{M}} |\log\cos\phi_1| \cos\phi d\phi dr =$$
$$= |\log C| + (2|\Gamma|)^{-1} \int_{\mathcal{M}} |\log\cos\phi| \cos\phi d\phi dr,$$

where at the last step we used the invariance of the measure $\mu$. Finally,

$$\int_{\mathcal{M}} |\log\cos\phi| \cos\phi d\phi dr = |\Gamma| \int_{-\pi/2}^{\pi/2} |\log\cos\phi| \cos\phi d\phi = |\Gamma|(2 - log4).$$

Hence,

$$\int_{\mathcal{M}} \log^+ ||\mathcal{D}_x\mathcal{F}|| \leq |\log C| + 1 - \log 2 < \infty.$$

$\square$

**Remark 4.3.8.** It is essential in the above proof that the curvature of $\Gamma$ is bounded. We used also the fact that $|\Gamma|$ is finite to define the probability measure $\mu$.

The preceding theorem leaves open the question of whether a given billiard is hyperbolic, i.e. has Lyapunov exponents almost everywhere positive. We will address this question very deeply in the next section, but before doing that let us see an example.

**Proposition 4.3.9.** *The Lyapunov exponents for the billiard map in a circle are zero at every point.*

*Proof.* Let us denote $\mathcal{F}_{circ}$ the billiard map for the unit circle billiard. The collision angle is equal to half of the angle correspondent to the arc between the two collision points and remains unchanged by the map, hence

$$\mathcal{F}_{circ}(r, \phi) = (r + 2\phi \pmod{2\pi}, \phi).$$

So that, for any $n \in \mathbb{N}$,

$$\mathcal{F}^n_{circ}(r, \phi) = (r + 2n\phi \pmod{2\pi}, \phi).$$

Differentiating the last equation, we obtain

$$\mathcal{D}_x\mathcal{F}_{circ} = \begin{bmatrix} 1 & 2n \\ 0 & 1, \end{bmatrix}$$

for any $x \in \mathcal{M}$. Thus the image of tangent vectors can grow at most linearly. $\square$

An analogue statement holds for a polygonal billiard.

**Proposition 4.3.10.** *The Lyapunov exponents for the billiard map in a polygon are zero at every point for which the asymptotic dynamics is defined.*

*Proof.* Let $\{A_n\}_{n\in\mathbb{N}^+}$ and $\{B_n\}_{n\in\mathbb{N}^+}$ two sequence of real numbers, we write $A_n \ll B_n$, $n \in \mathbb{N}^+$ if there exists a constant $D \in \mathbb{R}^+$ such that $A_n \leq DB_n$ for any $n$.
By (4.2.9) and the fact that $\mathcal{K}$ is zero everywhere, in that case

$$\mathcal{D}_x\mathcal{F}_{pol} = (-1) \begin{bmatrix} \frac{\cos\phi}{\cos\phi_1} & \frac{\tau_0}{\cos\phi_1} \\ 0 & 1 \end{bmatrix},$$

where $\tau_0 = \tau(x)$. Setting $\tau_i = \tau(\mathcal{F}^i(x))$ for any $i \in \mathbb{N}^+$ and applying the chain rule, we have

$$\mathcal{D}_x\mathcal{F}^n_{pol} = (-1)^n \begin{bmatrix} \frac{\cos\phi}{\cos\phi_n} & \sum_{i=0}^{n-1}(\frac{\tau_i}{\cos\phi_n}) \\ 0 & 1 \end{bmatrix}. \tag{4.3.16}$$

Furthermore, $\tau(x)$ is bounded by some constant $Q$. Therefore, we have that

$$\frac{1}{\cos\phi_n}\sum_{i=0}^{n-1}\tau_i \leq Q \times \frac{n}{\cos\phi_n}.$$

On the other hand, for almost every $x \in \mathcal{M}$ we have that

$$\left|\log(\frac{n}{\cos\phi_n})\right| \ll a \times n \quad ,$$

for any $a \in \mathbb{R}^+$. Indeed, this happens if and only if for almost every $x \in \mathcal{M}$

$$\left|\frac{\log n}{n} - \frac{\log(\cos\phi_n)}{n}\right| \ll a \quad .$$

Observe that $\frac{1}{n}\log\cos\phi_n \to 0$ almost everywhere (it is a simple consequence of the Birkhoff ergodic theorem that for any integrable function $g : \mathcal{M} \to \mathbb{R}$ we have $\frac{1}{n}g(\mathcal{F}^n x) \to 0$ almost everywhere). Hence by (4.3.16), at $\mu$-almost every point the iterated images of tangent vectors under $\mathcal{DF}$ grow asymptotically slower than any exponentially function of time $n$. $\square$

Before addressing the question of whether a map is hyperbolic in greater generality, we expose some more facts concerning Lyapunov exponents (about billiard dynamics but with wider applications). Since $\dim\mathcal{M} = 2$, we have, at every point $x \in \mathcal{M}$, either one Lyapunov exponent $\lambda$ of multiplicity two or two Lyapunov exponents $\lambda_x^{(1)} > \lambda_x^{(2)}$ of multiplicity one each.

**Lemma 4.3.11.** $\lambda_x^{(1)} + \lambda_x^{(2)} = 0$ *(up to a set of measure zero).*

*Proof.* Let $\Pi$ be a parallelogram in $\mathcal{T}_x\mathcal{M}$ with sides $a$ and $b$ parallel to $E_x^{(1)}$ and $E_x^{(2)}$, respectively. Its area is $\Pi = ab\sin\gamma$, where $\gamma$ denotes the angle between $E_x^{(1)}$ and $E_x^{(2)}$. Now $\mathcal{D}_x\mathcal{F}^n(\Pi)$ is a similar parallelogram in $\mathcal{T}_{\mathcal{F}^n x}\mathcal{M}$ whose sides we denote by $a_n$ and $b_n$, and the angle between them by $\gamma_n$. Therefore,

$$|\det \mathcal{D}_x\mathcal{F}^n| = \frac{\sin\gamma_n}{\sin\gamma}\frac{a_n}{a}\frac{b_n}{b}. \tag{4.3.17}$$

Due to the definition of Lyapunov exponents, we have

$$n^{-1}\log(a_n/a) \to \lambda_x^{(1)}, \quad n^{-1}\log(b_n/b) \to \lambda_x^{(2)}, \tag{4.3.18}$$

as $n \to +\infty$. On the other hand, by using (4.2.9) one obtains

$$|\det \mathcal{D}_x\mathcal{F}^n| = \frac{\cos\phi}{\cos\phi_n}, \tag{4.3.19}$$

where $\phi$ and $\phi_n$ are the $\phi$-coordinates of the points $x$ and $\mathcal{F}^n x$, respectively. Observe again that $\frac{1}{n}\log\cos\phi_n \to 0$ almost everywhere (see the proof of Proposition 4.3.10). Now taking the logarithms of (4.3.17) and (4.3.19) and dividing by $n$ we have

$$\frac{1}{n}\big(\log\sin\gamma_n - \log\sin\gamma + \log(a_n/a) + \log(b_n/b)\big) = \frac{1}{n}\big(\log\cos\phi - \log\cos\phi_n\big),$$

taking the limit $n \to \infty$ and using (4.3.18) and Proposition 4.3.4, we prove the lemma. □

We see that at almost every point $x \in \mathcal{M}$ the set of Lyapunov exponents (which is called *Lyapunov spectrum*) is symmetric about zero: either zero is the only Lyapunov exponent (of multiplicity two), or the two distinct Lyapunov exponents have equal absolute value but opposite signs. In the latter case $x$ is hyperbolic.

### 4.3.2 Proving hyperbolicity: cone techniques

We now discuss a general method for establishing hyperbolicity for phase space points $x \in \mathcal{M}$ and constructing their stable and unstable subspaces $E^u_x$ and $E^s_x$. We will use the billiard notation $\mathcal{F} : \mathcal{M} \to \mathcal{M}$, but the method applies without change to any type of maps introduced before as long as $\dim M = 2$ (but can also be generalized to higher dimensions). We put $x_n := \mathcal{F}^n x$ for all $n \in \mathbb{Z}$.
The basic idea of constructing the unstable space $E^u_x$ is to choose some non-zero vectors $v_{-n} \in \mathcal{T}_{x_{-n}}\mathcal{M}$ for $n \geq 0$ and then take the limit

$$E^u_x = \lim_{n\to\infty} span\{\mathcal{D}_{x_{-n}}\mathcal{F}^n(v_{-n})\}. \tag{4.3.20}$$

Of course, (4.3.20) need not hold for *all* choices of $\{v_n\}$, but it should hold for typical choices, as the following (semi-heuristic) argument suggests. Suppose the point $x$, and hence $x_n$ for all $n$, is indeed hyperbolic. Then $v_{-n} = c_{-n,u}v_{-n,u} + c_{-n,s}v_{-n,s}$ where $v_{-n,u} \in E^u_{x_{-n}}$ and $v_{-n,s} \in E^s_{x_{-n}}$ are some *unit* vectors. Due to (4.3.2) one expects that

$$\mathcal{D}_{x_{-n}}\mathcal{F}^n(v_{-n}) \sim e^{\lambda_x n}c_{-n,u}w_{-n,u} + e^{-\lambda_x n}c_{-n,s}w_{-n,s},$$

where $\lambda_x > 0$ is a positive Lyapunov exponent and $w_{n,u} \in E^u_x$ and $w_{n,s} \in E^s_x$ are some *unit* vectors. Now, if the coefficients $c_{-n,u}$ and $c_{-n,s}$ are bounded away from zero and infinity, i.e.

$$0 < c_{min} \leq |c_{-n,u}|, \quad |c_{-n,s}| \leq c_{max} < \infty,$$

then (4.3.20) obviously holds. In this case we can estimate the Lyapunov exponent $\lambda_x > 0$ by

$$\lambda_x = \lim_{n\to\infty}\frac{1}{n}\log||\mathcal{D}_{x_{-n}}\mathcal{F}^n(v_{-n})||/||v_{-n}||. \tag{4.3.21}$$

Furthermore, both (4.3.20) and (4.3.21) hold unless the ration $|c_{-n,s}/c_{-n,u}|$ grows, as $n \to \infty$ as fast as $e^{2\lambda_x n}$ or faster. Thus we need only to choose the initial vectors $v_{-n}$ not to close to the unstable spaces $E^s_{x_{-n}}$ (a very mild requirement, but it still needs to be met). To choose $v_{-n}$ properly one can rely on the method of cones. Suppose our dynamical system is endowed with a measure $\mu$.

**Definition 4.3.12.** Let $L \subset \mathcal{T}_x\mathcal{M}$ be a line and $\alpha \in (0, \pi/2)$. A *cone* $\mathcal{C}$ with axis $L$ and opening $\alpha$ is the set of all tangent vectors $v \in \mathcal{T}_x\mathcal{M}$ that make angle $\leq \alpha$ with the line $L$. The boundary $\partial\mathcal{C}$ consists of vectors making angle $\alpha$ with $L$ (this includes the zero vector). A *cone field* $\mathcal{C}^u(x)$ is the assignment to every (or $\mu$-almost every) point of a cone, in a way such that the functions $L(x)$ and $\alpha(x)$ defined everywhere (or $\mu$-almost everywhere) are measurable.

In his article [W1], Wojtkowski gives a sufficient condition for proving hyperbolicity entirely trough the invariance of cones. To present his result we define what is an *eventually strictly invariant cone field.* His ideas work for symplectic maps only, but in 2D that condition is equivalent to the preservation of an absolutely continuous measure. That is true for the billiard map for which the preserved measure is $\cos\phi d\phi dr$.

**Definition 4.3.13.** A cone filed $\mathcal{C}^u(x)$ is said invariant, if

$$\mathcal{D}_x\mathcal{F}(\mathcal{C}^u(x)) \subseteq \mathcal{C}(\mathcal{F}x),$$

for $\mu$-almost every $x \in \mathcal{M}$.

**Definition 4.3.14.** We say that a cone field $\mathcal{C}^u(x)$ is *eventually strictly invariant*, if it is invariant and, for $\mu$-almost every $x$, there exists an $n(x) \in \mathbb{N}^+$ such that

$$\mathcal{D}_x\mathcal{F}^{n(x)}(\mathcal{C}^u(x)) \subsetneq \mathcal{C}(\mathcal{F}^{n(x)}x).$$

**Theorem 4.3.15.** *Given a map $\mathcal{F}$ that preserves the absolutely continuous measure $\mu$, if there exists an eventually strictly invariant cone bundle, then the bigger Lyapunov exponent $\lambda_x^{(1)}$ is positive for $\mu$-almost every $x \in \mathcal{M}$ (i.e $\mathcal{F}$ is hyperbolic).*

The precedent theorem is very interesting because it gives a simple idea of an example of *non-uniformly* hyperbolic behavior. Indeed, the moment in which the real hyperbolic nature of a map (conserving an absolutely continuous measure) is displayed is precisely when it maps a cones *strictly* inside another. Suppose we look at the action of the differential $\mathcal{D}_x\mathcal{F}^n$ along the orbit $\{\mathcal{F}^n x\}_{n\geq 0}$ of $x$. Before seeing the cone $\mathcal{C}(\mathcal{F}^i x)$ being mapped *strictly inside* the cone $\mathcal{C}(\mathcal{F}^{i+1}x)$ one has to wait time $n(x)$, according to the definition of *eventually* strictly invariant cones. After that, one waits time $n(\mathcal{F}^{n(x)}x)$ before seeing the strictly invariance again, and so on. If for any constant $A \in \mathbb{R}^+$ the set of points $x \in \mathcal{M}$ such that $n(x) \geq A$ has positive measure, one does not have any upper bound on the waiting time for seeing the strictly invariance of cones, and this generally makes the hyperbolicity non-uniform.

Another way of characterizing hyperbolicity is trough invariance *and* expansion.

**Theorem 4.3.16.** *Assume that the cones $\mathcal{C}^u(x)$ satisfy*

$$\mathcal{D}_x\mathcal{F}(\mathcal{C}^u(x)) \subseteq \mathcal{C}^u(\mathcal{F}x) \quad \textit{(invariance)} \tag{4.3.22}$$

*and for all $v \in \mathcal{C}^u(x)$*

$$||\mathcal{D}_x\mathcal{F}(v)|| \geq \Lambda||v|| \quad \textit{(expansion).} \tag{4.3.23}$$

*Then the map $\mathcal{F}$ at almost every point $x \in \mathcal{M}$ has a positive Lyapunov exponent $\lambda_x \geq \log\Lambda$. Furthermore $E^u_x \subseteq \mathcal{C}^u(x)$ at almost every point and*

$$E^u_x = \bigcap_{n\geq 0} \mathcal{D}_{x_{-n}}\mathcal{F}^n(\mathcal{C}^u(x_{-n})). \tag{4.3.24}$$

*(This result justify (4.3.20) with any choice of $v_{-n} \in \mathcal{C}^u(x_{-n})$ ). An analogous statement holds for stable subspaces and cones substituting $\mathcal{F}$ with $\mathcal{F}^{-1}$.*

*Proof.* (Sketch) Let $\mathcal{T}_x\mathcal{M} = E^u_x \oplus E^s_x$ be the invariant split of the tangent space at (almost) any $x \in \mathcal{M}$ given by Oseledets theorem. For any $n \in \mathbb{N}^+$, pick any vector $v_{-n}$ in $\mathcal{C}^u(\mathcal{F}^{-n}x)$. Let us write $v_{-n} = c_{-n,u}v_{-n,u} + c_{-n,s}v_{-n,s}$ where $v_{-n,u}$ and $v_{-n,s}$ are unit vectors belonging to $E^u_{\mathcal{F}^{-n}x}$ and $E^s_{\mathcal{F}^{-n}x}$, respectively. If $E^u_{\mathcal{F}^{-n}x}$ were not inside $\mathcal{C}^u(\mathcal{F}^{-n}x)$, then, by the invariance of cones and of the splitting, $v := \mathcal{D}_{\mathcal{F}^{-n}x}\mathcal{F}^n(v_{-n})$ belongs to $\mathcal{C}(x)$ which does not contain $E^u_x$. Let us expand $v$ as $v = c_uv_u + c_sv_s$ where $v_u$ and $v_s$ are unit vectors belonging to $E^u_x$ and $E^s_x$, respectively. Since $E^u_x \notin \mathcal{C}(x)$, we have that $c_s/c_u > D$ for some positive constant $D$, for any $v \in \mathcal{C}^u(x)$ and hence for any choice of $v_n \in \mathcal{C}^u(\mathcal{F}^{-n}x)$. But in this way we arrive at a contradiction. Indeed, by (4.3.23) we have that the modulus of $v$ is grater with respect to the modulus of $v_{-n}$ by a factor $\Lambda^n$, i.e.,

$$||\mathcal{D}_{\mathcal{F}^{-n}x}\mathcal{F}^n v_{-n}|| \geq \Lambda^n||v_{-n}||. \tag{4.3.25}$$

On the other hand, if $v \in \mathcal{C}(x)$ and $E^u_x \notin \mathcal{C}(x)$ then one would have $c_s/c_u > D$ which is absurd because by (4.3.2) one has

$$\mathcal{D}_{\mathcal{F}^{-n}}\mathcal{F}^n(v_{-n}) \sim \underbrace{e^{\lambda_x n}c_{-n,u}}_{c_u}v_u + \underbrace{e^{-\lambda_x n}c_{-n,s}}_{c_s}v_s. \tag{4.3.26}$$

We remind that $\lambda^{(1)}_x = -\lambda^{(2)}_x$ under our assumptions (see Lemma 4.3.11). This argument proves that $E^u_x \subset \mathcal{C}^u(x)$ almost everywhere. Furthermore, from the fact that $E^u_{\mathcal{F}^{-n}x} \subset \mathcal{C}^u(\mathcal{F}^{-n}x)$ one has that $E^u_x \subset \bigcap_{n=0}^N \mathcal{D}_{x_{-n}}\mathcal{F}^n(\mathcal{C}^u(x_{-n})$ for any $N \in \mathbb{N}^+$ and (4.3.24) holds. The fact that $\lambda_x \geq \log\Lambda$ is clear once we compare (4.3.25) and (4.3.26). $\square$

Clearly, since for every $v \in E^u_x \subset \mathcal{C}^u(x)$ we have $||\mathcal{D}_x\mathcal{F}^n v|| \geq \Lambda^n ||v||$ the hyperbolicity is uniform under the condition of Theorem 4.3.16. It is interesting to note that none of the two assumptions of expansion and invariance alone can guarantee hyperbolicity.

- The identity map on any manifold $M$ leaves invariant any measurable cone field but clearly isn't hyperbolic.
- Consider the map $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\vec{x} \mapsto R \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \vec{x}$ for any $R > 1$ and $\theta \in (0, 2\pi)$. Being linear, it acts in the same way both in $\mathbb{R}^2$ and in $\mathcal{T}_{\vec{x}}\mathcal{M}$. We note that $R_\theta$ is expanding but does not leaves any cone invariant so that there cannot be an invariant splitting of of the tangent space.

### 4.3.3 Cones and geometrical optics

Now we give a incomplete and modest introduction to some recent developments in the study of hyperbolic billiards. In his beautiful paper of 1986 [W2], Wojkowski came up with a new method for addressing the problem of hyperbolicity which makes use of eventually strictly invariance of cones. His method allows us to prove eventually strictly invariance for hyperbolic billiards without doing directly computations with the differential of the map $\mathcal{DF}$; furthermore, it characterize hyperbolicity in a very visible fashion unrevealing to us a whole class of hyperbolic billiards. Suppose we want to prove hyperbolicity trough Theorem 4.3.15. The difference between Theorem 4.3.15 and Theorem 4.3.16 is that the first does not care about the *increasing magnitude* of images of vectors under $\mathcal{DF}$, while the second does ask expansion. Roughly speaking, Theorem 4.3.15 requires that we know only how the *inclination* of repeated images of vectors change, to see whether they stay inside some cones in the tangent space. The main idea is to study the action of the differential $\mathcal{D}_x\mathcal{F}$ only on the inclination of vectors belonging to $\mathcal{T}_x\mathcal{M}$ (which is a *projective coordinate*).

- **Focusing times.** We introduce focusing times for infinitesimal families of trajectories, which will be used as projective coordinates.
  A *variation* $\eta(\alpha)$ is a one-parameter smooth family of lines in $\mathbb{R}^2$. Let $z \in \mathcal{M}$, and $\xi : (-\epsilon, \epsilon) \to \mathcal{M}$ be a curve such that $\xi(0) = z$ and $\xi'(0) = u \in \mathcal{T}_z\mathcal{M}$. We associate with $u$ the variation

  $$\eta_+(\alpha) = \{q(\alpha) + tv(\alpha), t \in \mathbb{R}, |\alpha| < \epsilon\},$$

  where $\xi(\alpha) = (q(\alpha), v(\alpha))$, $q(\alpha) \in \partial\mathcal{D}$. We say that $\eta_+(\alpha)$ *focuses at time* $f^+$ if there exists $f^+ = \lim_{\alpha\to 0} t(\alpha)$, where $t(\alpha)$ is determined by the intersection of the lines $\eta_+(\alpha)$ and $\eta_+(0) = q(0) + tv(0)$. Also we consider the variation

  $$\eta_-(\alpha) = q(\alpha) + t\tilde{v}(\alpha), \quad t \in \mathbb{R}$$

where the vector $\tilde{v}(\alpha)$ is obtained by reflecting $v(\alpha)$ across the line tangent to $\partial\mathcal{D}$ at the point $q(\alpha) \in \partial\mathcal{D}$. We define the *focusing* time $f^-$ of the variation $\eta_-(\alpha)$ in the same way as before. Let $\eta_\pm(\alpha, t)$ be the parametric representation of $\eta_\pm(\alpha)$

$$\eta_\pm(\alpha, t) = \{q(\alpha) + tv(\alpha), |\alpha| < \epsilon\} \quad t \in \mathbb{R}.$$

Although for each vector $u \in \mathcal{T}_z\mathcal{M}$ we can construct infinitely many distinct variations $\eta_+(\alpha)(\eta_-(\alpha))$, all of which will focus at the same time $f^+$ (respectively, $f^-$). An alternative definition of the focusing time is the following: $\eta_+(\alpha)$ focuses if the vector $\partial\eta_+(\alpha, t(\alpha))/\partial\alpha_{|\alpha=0}$ is parallel to $v(0)$ for some $t = f^+ \in \mathbb{R}$. We say that the vector $u$ focuses forward (backward) if the variation $\eta_+(\alpha)$ (respectively, $\eta_-(\alpha)$) focuses.
Consider the two vector fields $X_r = \partial/\partial r$ and $X_\phi = \partial/\partial\phi$ on $\mathcal{M}$. Given a point $z = (r, \phi) \in \mathcal{M}$ and a vector $u = u_r\partial/\partial r + u_\phi\partial/\partial\phi \in \mathcal{T}_z\mathcal{M}$ with $u_r, u_\phi \in \mathbb{R}$, the *forward* and *backward focusing times* $f^+(u), f^-(u)$ of $u$ are given by (for the derivation see [W2])

$$f^\pm(u) = \begin{cases} \frac{-\cos\phi}{\mathcal{K}(r)\pm m(u)}, & \text{if } u_r \neq 0 \\ 0 & \text{if } u_r = 0 \end{cases}$$

where $m(u) = u_\phi/u_r$ is the slope of $u$.
We say that $u$ is *divergent* (*convergent*) if $f^+(u)$ is negative (positive) and $u$ is *flat* if $f^+(u) = \infty$. Later on, when we will consider cone fields, we will use the notation $f^\pm(z, u)$, rather than $f^\pm(u)$, to indicate the dependence on the point $z$.

- **Reflection Law.** Let $z = (r, \phi) \in \mathcal{M} \setminus \mathcal{S}_1$ so that the point $\mathcal{F}z = (r_1, \phi_1)$ is well defined. For any $u \in \mathcal{T}_z\mathcal{M}$, let $f_0 = f^+(u)$ and $f_1 = f^+(\mathcal{D}_z\mathcal{F}u)$. Then one can prove that the action of $\mathcal{DF}$ is reflected on the projective coordinates $f_0$ and $f_1$ in the *mirror equation* which takes the form

$$\frac{1}{f_1} + \frac{1}{\tau(z) - f_0} = \frac{-2\mathcal{K}(r_1)}{\cos\phi_1} \tag{4.3.27}$$

Under suitable conditions, the focusing time $f$ has an important ordering property that will be crucial for the proof of hyperbolicity of billiard maps.

**Lemma 4.3.17.** *(Ordering property). Let $z \in \mathcal{M} \setminus \mathcal{S}_1$ and $u, v \in \mathcal{T}_z\mathcal{M}$. Assume that $0 < f^+(z, w) < \tau(z)$ and $0 < f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}w)$. Then*

$$0 \leq f^+(z, u) \leq f^+(z, w) \Rightarrow 0 < f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}u) \leq f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}w).$$

*The implication is also true if we replace the inequalities with strict inequalities.*

*Proof.* By formula (4.3.27), we have that $1/f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}w) < 1/f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}u)$. $\square$

One can see immediately that $f^+$ is a projective coordinate in $\mathcal{T}_z\mathcal{M}$. A cone $\mathcal{C}(z) \subset \mathcal{T}_z\mathcal{M}$ can be identified with a closed interval of values of $f^+$:

$$\mathcal{C}(z) = \{u \in \mathcal{T}_z\mathcal{M} : f_1 \leq f^+(u) \leq f_2\}$$

for some real numbers $f_1, f_2$ such that $-\infty \leq f_1 < f_2 \leq +\infty$.

**Definition 4.3.18.** Let $\mathcal{C}(z)$ be a cone field on $\mathcal{M}_-$ and $\mathcal{C}'(z)$ be the complementary cone field. For every $z \in \mathcal{M}_-$, define

$$f^+(z) = \sup_{u \in \mathcal{C}(z)} f^+(z,u), \quad f^-(z) = \sup_{u \in \mathcal{C}'(z)} f^-(z,u).$$

With this machinery at our disposal, one can prove easily that a table whose boundary is composed by dispersing wall only, i.e. $\mathcal{M} = \mathcal{M}_+$ is hyperbolic.

**Proposition 4.3.19.** *If $\mathcal{M} = \mathcal{M}_+$, then the billiard map $\mathcal{M}$ is hyperbolic.*

*Proof.* Let us define the following cone field

$$\mathcal{C}^u(z) = \{u \in \mathcal{T}_z\mathcal{M} : f^+(z,u) \leq 0\}. \tag{4.3.28}$$

By Equation (4.3.27), one has that

$$f^+(z,u) < 0 \Rightarrow f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}u) < 0,$$

and

$$f^+(z,u) = 0 \Rightarrow f^+(\mathcal{F}z, \mathcal{D}_z\mathcal{F}u) < 0.$$

In the last equation we used the fact that $\mathcal{M} = \mathcal{M}_+$ and so $\mathcal{K} > 0$ everywhere. Therefore the cone $\mathcal{C}^u(z)$ is strictly invariant at every iteration of the differential. □

The next result gives a simple criterion which helps to check whether a cone field on $\mathcal{M}_-$ (which is the subset of the phase space correspondent to negative (focusing) curvature) is invariant. It was first proved in [D].

**Lemma 4.3.20.** *(Focusing Lemma). Let $z \in \mathcal{M}_- \setminus \mathcal{S}_1$ such that $\mathcal{F}z \in \mathcal{M}_-$. Suppose that $0 \leq f^+(z,u) \leq \tau(z)$ for every $u \in \mathcal{C}(z)$ and $0 < f^-(\mathcal{F}z) \leq \tau(z)$. Then*

$$0 < f^+(z) + f^-(\mathcal{F}z) \leq \tau(z) \Rightarrow \mathcal{D}_z\mathcal{F}\mathcal{C}(z) \subseteq \mathcal{C}(\mathcal{F}z),$$

*and if the inequality on the left-hand side is strict, then we get strict invariance: $\mathcal{D}_z\mathcal{F}\mathcal{C}(z) \subset int\mathcal{C}(\mathcal{F}z)$.*

*Proof.* Let $w_1 \in \mathcal{C}(\mathcal{F}z)$ such that $f^+(\mathcal{F}z, w_1) = f^+(\mathcal{F}z)$; in other words $w_1$ belongs to one of the edges of $\mathcal{C}(\mathcal{F}z)$. If $w = \mathcal{D}_{\mathcal{F}z}\mathcal{F}^{-1}w_1$, then $f^+(z, w) = \tau - f^-(\mathcal{F}z)$. Now if $u \in \mathcal{C}(z)$, then $f^+(z, u) \leq f^+(z)$. By the hypothesis we have $f^+(z) \leq (<)\tau(z) - f^-(\mathcal{F}z) = f^+(z, w)$ so that $f^+(z, u) \leq (<) f^+(z, w)$. To complete the proof one applies Lemma 4.3.17. □

The above considerations (especially Lemma 4.3.20) led to the discovery of a whole new class of hyperbolic non-dispersing billiards. The general idea is that also a focusing billiard can be hyperbolic as long as the focusing components (a) have boundaries described by some particular class of functions (b) are far enough one from the each other. Hyperbolicity comes from the fact that focusing rays, after a focusing collision, have enough time to focus and then defocus again before the next collision with the wall, thus yielding the same effect of a 'net' dispersing collision. In the literature, such a phenomenon is often referred to as the *Wojtkowski–Markarian–Donnay–Bunimovich technique for the hyperbolicity of focusing billiard.* There are also known example of non-standard focusing hyperbolic billiards which do not fulfill condition (b); i.e. they do not have the focusing components far enough considering their (*small*) radii of curvature. This class of hyperbolic billiards have been introduced by Lenci and Bussolari in [BL].

## 4.4 LSUM: the role of singularities

To explain better what LSUM are it is convenient to think of the concrete example of *dispersing billiards.*

**Definition 4.4.1.** A billiard table $\mathcal{D} \subset \mathbb{R}^2$ (or $\mathcal{D} \subset \mathbb{T}^2$) is said to be *dispersing* if all the walls $\Gamma_i \subset \Gamma = \partial\mathcal{D}$ are dispersing, i.e. $\Gamma = \Gamma_+$.

One can further classify dispersing billiards according to the following attributes:

- whether or not $\mathcal{D}$ has corners; if yes, whether it has cusps;
- if $\mathcal{D} \subsetneq \mathbb{T}^2$, whether or not the horizon is bounded.

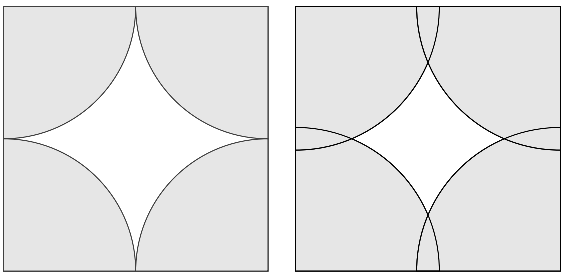

Figure 6 shows a finite-horizon billiard with cusps (left) and a finite horizon billiard with vertices and no cusps (right).

To simplify our analysis we will consider only billiards with no corners and bounded horizon, which we will call billiards of category (A) adopting the terminology of [CM].

### 4.4.1 Hyperbolicity

As a first step, let us establish (uniform) hyperbolicity for billiards of type (A).

**Proposition 4.4.2.** *Any dispersing billiard of category (A) is uniformly hyperbolic.*

*Proof.* We have already proved in Proposition 4.3.19 that $\mathcal{F}$ is hyperbolic. Put $\mathcal{V} = \frac{d\phi}{dr}$. To prove the uniform hyperbolicity we define here the *narrow unstable cones*

$$\hat{\mathcal{C}}^u_x = \{u = (dr, d\phi) \in \mathcal{T}_x\mathcal{M} : \mathcal{K}(x) \leq \mathcal{V} \leq \mathcal{K}(x) + \cos\phi(x)/\tau(\mathcal{F}^{-1}x)\}. \tag{4.4.1}$$

They are called narrow because they lie strctly inside the standard unstable cones $\mathcal{C}^u_x$ introduced before (see Equation (4.3.28)), i.e. $\hat{\mathcal{C}}^u_x \subsetneq \mathcal{C}^u_x$. By a direct calculation using (4.2.9) one can prove that $\hat{\mathcal{C}}^u_x$ are strictly invariant

$$\mathcal{D}_x\mathcal{F}(\hat{\mathcal{C}}^u_x) \subsetneq \hat{\mathcal{C}}^u_{\mathcal{F}x}.$$

For billiards of category (A), one has that there exist $\mathcal{V}_{min}$ and $\mathcal{V}_{max}$ such that if $(dr, d\phi) \in \hat{\mathcal{C}}^u_x$ then $\mathcal{V}_{min} \leq \mathcal{V} \leq \mathcal{V}_{max}$. Indeed, by the fact that the billiard is dispersing and assumption (A3)

$$0 < \mathcal{K}_{min} \leq \mathcal{K}(x) \leq \mathcal{K}_{max} < \infty.$$

Furthermore, since there are no corners and cusps and the billiard has finite horizon

$$0 < \tau_{min} \leq \tau(x) \leq \tau_{max} < \infty.$$

Let $dx = (dr, d\phi) \in \hat{\mathcal{C}}^u_x$ be an unstable vector, $dx_n = (dr_n, d\phi_n) = \mathcal{D}_x\mathcal{F}(dx)$ its image and $\tau_n = \tau(\mathcal{F}^n x)$. Using the expression (4.2.9) for the differential of the map and setting $\mathcal{V}_n = d\phi_n/dr_n$, we have that the expansion factor of $dx$ is given by

$$\frac{||dx_{n+1}||}{||dx_n||} = (1 + \tau_n(\mathcal{V}_n + \mathcal{K})/\cos\phi_n)\frac{\cos\phi_n}{\cos\phi_{n+1}}\frac{\sqrt{1+\mathcal{V}^2_{n+1}}}{\sqrt{1+\mathcal{V}^2_n}}$$

Therefore

$$\frac{||dx_n||}{||dx_0||} = \prod_{i=0}^{n-1}(1 + \tau_i(\mathcal{V}_i + \mathcal{K})/\cos\phi_i)\frac{\cos\phi_0}{\cos\phi_n}\frac{\sqrt{1+\mathcal{V}^2_n}}{\sqrt{1+\mathcal{V}^2_0}}.$$

Since we are considering billiards with no corners or cusps then $\tau(x) \geq \tau_{min}$ for every $x \in \mathcal{M}$ and some $\tau_{min} \in \mathbb{R}^+$. Furthermore $\mathcal{V} \geq \mathcal{K}(x)$ on the unstable cones and $\mathcal{K} \geq \mathcal{K}_{min} > 0$ since the billiard is dispersing. This means that if we set

$$\mathcal{R}_{min} := 2\mathcal{K}_{min} > 0^6$$

[6] $\mathcal{R} := 2\mathcal{K}/\cos\phi > \mathcal{R}_{min}$ is often called collision parameter since it describes how it changes the curvature of an infinitesimal beam of trajectories after a collision.

then

$$\frac{||dx_n||}{||dx_0||} \geq (1 + \tau_{min}\mathcal{R}_{min})^n \frac{\cos\phi_0}{\cos\phi_n} \frac{\sqrt{1+\mathcal{V}_n^2}}{\sqrt{1+\mathcal{V}_0^2}}.$$

Setting

$$(1 + \tau_{min}\mathcal{R}_{min}) := \Lambda > 1$$

one has

$$\frac{||dx_n||}{||dx_0||} \geq \Lambda^n \frac{\cos\phi_0}{\cos\phi_n} \frac{\sqrt{1+\mathcal{V}_n^2}}{\sqrt{1+\mathcal{V}_0^2}}.$$

Now, since we are $dx \in \hat{\mathcal{C}}_x^u$ one has that $\mathcal{V}_n \geq \mathcal{K}_{min} > 0$ for all $n \in \mathbb{N}^+$ so that there exists $\hat{c} \in \mathbb{R}^+$ such that

$$\frac{||dx_n||}{||dx_0||} \geq \hat{c}\Lambda^n, \tag{4.4.2}$$

and uniform hyperbolicity is established. □

If one finds the previous calculation a bit involved, it's because it is. Actually, hidden in the previous proof, there is a much more simple interpretation of uniform expansion of vectors of unstable cones. Let us introduce the *p-metric* defined by

$$||dx||_p = |\cos\phi| dr$$

for any vector $(dr, d\phi) \in \mathcal{T}_x\mathcal{M}$. What does this norm represent? We have seen when we talked about geometric optics that we can describe vectors (actually, their slope) in $\mathcal{T}_x\mathcal{M}$ using infinitesimal beam of trajectories. Indeed, any vector $dx \in \mathcal{T}_x\mathcal{M}$ is naturally associated with a differentiable curve $\phi : (-\epsilon, \epsilon) \to \mathcal{M}$ such that $\phi(0) = x$ and $\phi'(0) = v$. To this function we can associate what we called a variation, i.e. if $\phi(\sigma) = (r(\sigma), v(\sigma))$ with $r(\sigma) \in \partial\mathcal{D}$ and $v(\sigma) \in S^1$ then the associated variation is $\eta^+ = \{r(\sigma) + rv(\sigma) | r \in \mathbb{R}, -\epsilon \leq \sigma \leq \epsilon\}$. A variation is an infinitesimal beam of trajectories. The fact that $dx \in \hat{C}_x^u$ means that $d\phi/dr > 0$ and so the beam will have positive curvature (we say that it is *divergent*). The norm $||dx||_p$ represents (infinitesimally) the portion of the walls on which the infinitesimal beam associated to $dx$ projects. If we had studied the action of the differential $\mathcal{D}_x\mathcal{F}$ on the vector $dx$ with respect to the p-norm, we would have found the easier equation

$$\frac{||dx||_{p,n+1}}{||dx||_{p.n}} = (1 + \tau_n(\mathcal{V}_n + \mathcal{K})/\cos\phi_n) \geq (1 + \tau_{min}\mathcal{R}_{min}) \geq \Lambda.$$

Visualizing what is happening is now possible. A vector in $\hat{C}_x^u$ is associated with a divergent beam. The fact that $\hat{C}_x^u$ is invariant mirrors the fact that divergent beams stay divergent when colliding with dispersing boundaries. Furthermore, the fact that the p-norm increases means that, infinitesimally, the beam projects on larger and larger portions of the boundary $\partial\mathcal{D}$ collision after collision. This happens because:

a) once we star with a vector in $\hat{C}^u_x$, the incident beam, which starts as divergent, remains always divergent (the cone is invariant!);

b) a divergent beam has the property that trajectories 'separates one from the other' on their trajectory. The longer is the trajectory, the greater is the separation between 'infinitesimally near orbits' between two collision. If the return time is bounded from below then this separation grows of a minimal amount at every collision and the projection of the divergent beam on the boundary of the table grows every time.

This way of thinking is certainly nicer than doing the calculations of Proposition 4.4.2, but it proves the expansion of the p-metric not of our Euclidean metric $||dx|| = \sqrt{dr^2 + d\phi^2}$. Nonetheless, one can prove that these two metric are actually equivalent on the unstable cones (but is what we need). The p-metric has also the advantage of being monotone, i.e.

$$||\mathcal{D}_x\mathcal{F}(dx)||_p \geq ||dx||_p,$$

always. This may not be true for the Euclidean metric. Even though we have expansion, it may happen that the norms actually decrease for the first iterations of $\mathcal{DF}$ before starting to grow. Actually, the existence of an increasing metric is no matter of luck. It can be proved that (even for non uniformly) hyperbolic systems there always exists one, which is often called *adapted or Lyapunov metric* (see [P]).

Now we introduce stable and unstable curves that play an instrumental role in the studies of ergodic and statistical roles of dispersing billiards.

**Definition 4.4.3.** Suppose a curve $W \subset \mathcal{M}$ is defined by a function $\phi = \phi_W(r)$ for some $r'_W \leq r \leq r''_W$. We say that $W$ is $C^m$ smooth ($m \geq 1$) if the function $\phi_W(r)$ is $C^m$ smooth, up to the endpoints $r'_W$ and $r''_W$. We will say that a $C^m$ smooth curve is *unstable* (or *increasing*) if at every point $x \in W$ the tangent line $\mathcal{T}_xW$ belongs to the unstable cone $\hat{C}^u_x$ (see (4.4.1)).

As we said above, a line in $\mathcal{T}_x\mathcal{M}$ is in correspondence with an infinitesimal curve which is in turn in correspondence with a variation, i.e. an infinitesimal beam of trajectories. If the vector belongs to the narrow cones $\hat{C}^u_x$ then the infinitesimal beam is dispersing. If instead of a vector we take an unstable curve, we have a whole wave front rather than an infinitesimal beam on the physical space. If the curve is unstable, then the wave front is dispersing, i.e. it will have negative curvature everywhere. Since dispersing wave front, upon colliding on dispersing walls, remain dispersing then we see that the image of an unstable curve is another unstable curve. Yet another way to see it is in the proof of the following theorem.

**Proposition 4.4.4.** *Let $W \subset \mathcal{M}$ be an unstable curve. For billiards of category (A) $\mathcal{F}^n(W)$ consists of a finite number of unstable curves.*

*Proof.* Let $dx \in \mathcal{T}_x W$ be a tangent vector to the unstable curve $W$. Then the tangent vector to $\mathcal{F}(W)$, where is defined, is $\mathcal{D}_x\mathcal{F}(dx)$. Since, by definition of unstable curve $dx \in \hat{\mathcal{C}}^u_x$ and by the invariance of the narrow cones, we have that $\mathcal{D}_x\mathcal{F}(dx) \in \hat{\mathcal{C}}^u_x$. By induction $\mathcal{D}_x\mathcal{F}^n(dx) \in \hat{\mathcal{C}}^u_x$ for every $n \geq 1$. Consider now the set $\mathcal{F}(W)$. Since the horizon is bounded, every wave front projects on the boundary in a finite union of smooth wave fronts. Indeed, $\mathcal{F}(W)$ projects on each wall $\Gamma_i$ only once, and there are a finite number of walls by hypothesis. Therefore $\mathcal{F}(W)$ consists of a finite union of smooth curves[7]. By induction, also $\mathcal{F}^n(W)$ is a union of smooth curves, for every $n \in \mathbb{N}^+$. □

### 4.4.2 Singularities

Recall the basic definitions

$$\mathcal{S}_0 := \partial\mathcal{M} = \{|\phi| = \pi/2\} \cup \Big(\cup_i(\{r = a_i\} \cup \{r = b_i\})\Big),$$

$$\mathcal{S}_1 := \mathcal{S}_0 \cup \{x \in int\mathcal{M} : \mathcal{F}(x) \notin int\mathcal{M}\} = \partial\mathcal{M} \cup \mathcal{F}^{-1}(\partial\mathcal{M}),$$
$$\mathcal{S}_{-1} := \mathcal{S}_0 \cup \{x \in int\mathcal{M} : \mathcal{F}^{-1}(x) \notin \mathcal{M}\} = \partial\mathcal{M} \cup \mathcal{F}(\partial\mathcal{M}).$$

We define, inductively,

$$\mathcal{S}_{n+1} = \mathcal{S}_n \cup \mathcal{F}^{-1}(\mathcal{S}_n) \quad \mathcal{S}_{-(n+1)} = \mathcal{S}_{-n} \cup \mathcal{F}(\mathcal{S}_{-n}).$$

It is easy to see that $\mathcal{S}_{n+1}$ and $\mathcal{S}_{-(n+1)}$ are the singularity sets for the maps $\mathcal{F}^{n+1}$ and $\mathcal{F}^{-(n+1)}$, respectively. We subdivide our analysis of singularities in a local analysis and in a global one.

#### Local analysis

Let us start with a consequence of the time reversibility of billiard dynamics.

**Proposition 4.4.5.** *If $(r, \phi) \in \mathcal{S}_n$, then $(r, -\phi) \in \mathcal{S}_{-n}$, i.e. the future and past singularities are specular with respect to the axis $\phi = 0$.*

*Proof.* By the involution property

$$\mathcal{F}^{-k} \circ \mathcal{I} = \mathcal{I} \circ \mathcal{F}^k, \quad k \in \mathbb{Z},$$

and the fact that $\mathcal{I}(\mathcal{S}_0) = \mathcal{S}_0$ one readily obtains that

$$\mathcal{I}(\mathcal{S}_n) = \bigcup_{i=0}^{n} \mathcal{I}(\mathcal{F}^i(\mathcal{S}_0)) = \bigcup_{i=0}^{n} \mathcal{F}^{-i}(\mathcal{I}(\mathcal{S}_0)) = \bigcup_{i=0}^{n} \mathcal{F}^{-i}(\mathcal{S}_0) = \mathcal{S}_{-n}.$$

□

[7] If the billiard has unbounded horizon such a union may be countable and infinite.

This symmetry allows us to restrict ourselves to the case $n < 0$; i.e. we consider the singularities of the past iterations of the map $\mathcal{F}$.

**Proposition 4.4.6.** *For each $n > 0$, the set $\mathcal{S}_{-n} \setminus \mathcal{S}_0$ consists of $C^1$ smooth unstable curves $S \subset \mathcal{M}$ (i.e., the tangent line $S$ at every point $x \in S$ belongs in the unstable cones $\hat{\mathcal{C}}^u_x$).*

*Proof.* It is enough to check this for $n = 1$ and then use the Proposition 4.4.4. Every curve $S \subset \mathcal{S}_{-1} \setminus \mathcal{S}_0$ is made by trajectories either (i) coming from grazing collisions or (ii) emanating from corners. Both types of trajectories obviously make dispersing wave fronts when they collide with $\partial\mathcal{D}$ and form the curve $S$. For proving $C^1$ smoothness we refer the interested reader to Proposition 4.41 of [CM]. □

We will sharpen the above proposition by showing that the slope of singularity curves $S \subseteq \mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1}$ approaches, as $n$ increases, the slope of the unstable space $E^x_u$ at nearby points $x$.

**Proposition 4.4.7.** *Let $\mathcal{D}$ be a billiard table of category (A). Then there exists a sequence $\beta_n \to 0$ as $n \to \infty$ such that for any curve $S \subseteq \mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1}$ and any point $y \in S$ there is an open neighborhood $U_y \subset \mathcal{M}$ such that the slope $\mathcal{V}_S(y)$ of the curve $S$ at $y$ satisfies*

$$\sup_{x \in U_y} |\mathcal{V}^u_x - \mathcal{V}_S(y)| < \beta_n,$$

*where $\mathcal{V}^u_x$ is the slope of the unstable space $E^u_x$ in $x$, i.e. $\mathcal{V}^u_x = d\phi_u/dr_u$ for any $(dr_u, d\phi_u) \in E^u_x$. (Of course, the supremum here is taken over $x \in U_y$ where $\mathcal{V}^u_x$ is defined.)*

*Proof.* (General idea) Here we give only some heuristic arguments without actually proving the theorem, we refer the interested reader to Proposition 4.42 of [CM]. By Proposition 4.4.6 one has that $\mathcal{S}_{-1} \setminus \mathcal{S}_0$ is made of unstable curves. Consider a curve $W'$ in $\mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1}$. By definition, there must exists a (possibly very small) piece of curve, call it $W$, belonging to $\mathcal{S}_{-1} \setminus \mathcal{S}_0$ such that $W' = \mathcal{F}^{n-1}W$. Let $W$ be one of such curves. Pick a vector $v \in \mathcal{T}_xW$. Let $x \in W$ and $x_{n-1} = \mathcal{F}^{n-1}x \in W'$. Furthermore, let $L$ is the line in $\mathcal{T}_x\mathcal{M}$ correspondent to the slope $\mathcal{V}_W(x)$ of the curve $W$ at $x$ and $L'$ is the line in $\mathcal{T}_{x_{n-1}}\mathcal{M}$ correspondent to the slope $\mathcal{V}_{W'}(x)$ of the curve $W'$. The relation between $L$ and $L'$ is that

$$\mathcal{D}_x\mathcal{F}^{n-1}L = L'.$$

On the other hand, the unstable space $E^u_{x_{n-1}}$ (supposing that it is defined) is given by pushforwards of unstable cones in the far past, i.e.

$$E^u_{x_{n-1}} = \bigcap_{i=0}^{\infty} \mathcal{D}_{\mathcal{F}^{-i}(x_{n-1})}\mathcal{F}^i(\mathcal{C}^u_{\mathcal{F}^{-i}(x_{n-1})}). \tag{4.4.3}$$

Now, by Proposition 4.4.6 one has that $L \subset \mathcal{C}^u_x = \mathcal{C}^u_{\mathcal{F}^{-n+1}x_{n-1}}$. Therefore, by (4.4.3) $L'$ will be a better and better approximation of $E_{x_{n-1}}$ for increasing $n$. □

Loosely speaking, Proposition 4.4.7 says that the singularity curves $S \subseteq \mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1}$ align with the unstable space $E^u$ as $n$ increases. This property is called *alignment of singularity lines.*

### Global analysis

**Proposition 4.4.8.** *For dispersing billiards with finite horizon (so category (A) is included), the set $\mathcal{S}_{-1} \setminus \mathcal{S}_0$ consists of finitely many compact*[8] *smooth unstable (increasing) curves, and the set $\mathcal{S}_1 \setminus \mathcal{S}_0$ consists of finitely many compact smooth stable (decreasing) curves.*
*For billiard without bounded horizon, the same holds, except that the number of the above curves may be infinite.*

*Proof.* For the finite horizon case the argument is the same as the one in the proof of Proposition 4.4.6: the curves in $\mathcal{S}_{-1} \setminus \mathcal{S}_0$ are traces left on $\mathcal{M}$ by finitely many dispersing wave fronts (those emanating from corner points and those formed by grazing collisions with walls). Now, if $\gamma$ is a dispersing wave front, its trace on $\mathcal{M}$ is a *finite* union of smooth curves if the horizon is bounded, and possibly infinite if the horizon is unbounded. Indeed, only in the latter case the same dispersing curve can project on a single wall multiple times. The proof for curves in $\mathcal{S}_1 \setminus \mathcal{S}_0$ is just the time reversal. □

Using mathematical induction, we can actually say that Proposition 4.4.8 extends without changes for the sets $\mathcal{S}_{-n} \setminus \mathcal{S}_0$ and $\mathcal{S}_n \setminus \mathcal{S}_0$ for all $n \geq 1$. Indeed, since curves of $\mathcal{S}_1 \setminus \mathcal{S}_0$ are smooth (up to some degree) and have tangent vectors lying in stable cones, they intersect each curve in $\mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1}$ in finitely many (or countably many, if the horizon is unbounded) points; therefore $\mathcal{S}_{-n-1} \setminus \mathcal{S}_{-n} = \mathcal{F}(\mathcal{S}_{-n} \setminus \mathcal{S}_{-n+1})$ will be a finite (or countable) union of smooth unstable curves.
The last property we study is often referred to as *continuation of singularity lines*; one can continue every smooth curve $S \subset \mathcal{S}_n \setminus \mathcal{S}_0$ monotonically up to the natural boundary of $\mathcal{M}$.

**Proposition 4.4.9.** *For each $n \neq 0$ every curve $S \subseteq \mathcal{S}_n \setminus \mathcal{S}_0$ is a part of some monotonic continuous (and piece-wise smooth) curve $\tilde{S} \subseteq \mathcal{S}_n \setminus \mathcal{S}_0$ wich terminates on $\mathcal{S}_0 = \partial\mathcal{M}$ (the curve $\tilde{S}$ is monotonically increasing for $n < 0$ and monotonically decreasing for $n > 0$).*

*Proof.* First, we check that the statement is true for $n = \pm 1$. Consider $n = -1$, the other case being just the time reversal situation. Let us adopt the same terminology of Proposition 4.4.6. Every singularity curve $S \subseteq \mathcal{S}_{-1} \setminus \mathcal{S}_0$ is made by: either a trajectories coming from a grazing collision (call it type (i) singularity curve) or trajectories coming from corner (call it type (ii)).

[8] More precisely, their closure is compact; the removal of $\mathcal{S}_0$ deprives some curves in $\mathcal{S}_{-1}$ of their endpoints.

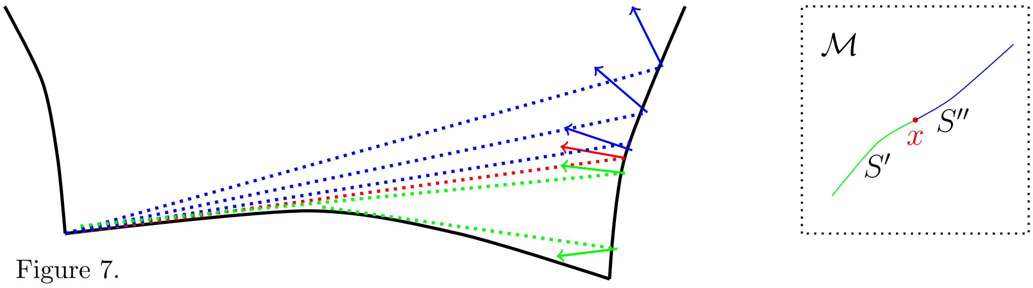


Figure 7.

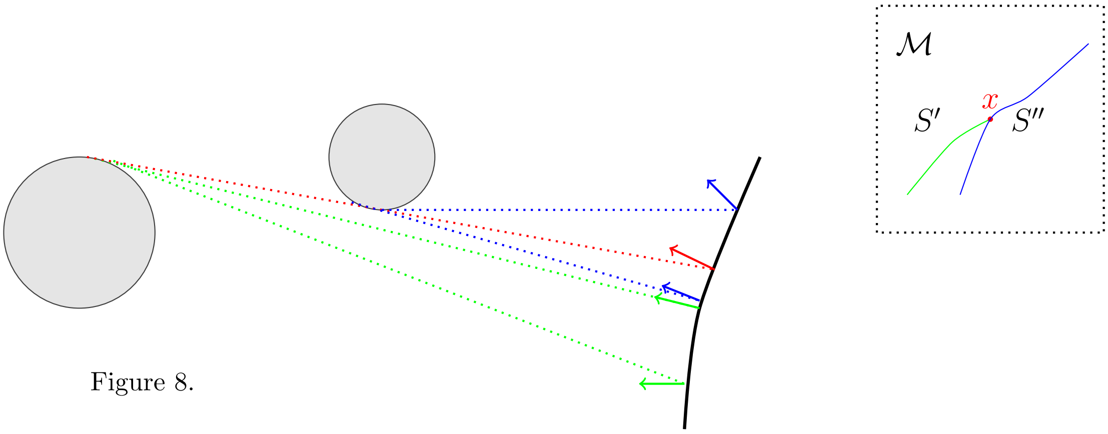


Figure 8.

Figure 7 shows a curve $S' \subseteq \mathcal{S}_{-1} \setminus \mathcal{S}_0$ of type (i) (induced by grazing collisions) terminating on a curve $S'' \subseteq \mathcal{S}_{-1} \setminus \mathcal{S}_0$ of type (ii) (coming from a corner point). Observe that the curves $S'$ and $S''$ have a common endpoint $x$ (and, furthermore, one can show that they have the same slope $\mathcal{V}(x)$, but their curvature at $x$ may differ). Figure 8 shows a curve $S' \subseteq \mathcal{S}_{-1} \setminus \mathcal{S}_0$ of type (i), which is induced by a grazing collision with scatterer 1, terminating on another, larger curve $S'' \subseteq \mathcal{S}_{-1} \setminus \mathcal{S}_0$ of type (i), which is induced by grazing collisions with scatterer 2. In both cases $S''$ terminates on another singularity curve. Another possibility is that the singular trajectories meet a vertex or a tangency to the boundary, i.e. $S''$ terminates on $\partial\mathcal{M}$. There are other few cases to consider but they are similar to the precedent two. Now that we have proved the statement for curves in $\mathcal{S}_{-1} \setminus \mathcal{S}_0$ we use induction. Suppose that the theorem holds for all the singularity curves in $\mathcal{S}_{-n} \setminus \mathcal{S}_0$. Then since by the inductive assumption each curve in $\mathcal{S}_{-n} \setminus \mathcal{S}_0$ is part of some monotonic continuous curve which terminates on the boundary $\partial\mathcal{M}$, it must be that each curve $S$ in $\mathcal{F}(\mathcal{S}_{-n} \setminus \mathcal{S}_0)$ is part of some monotonic curve $S'$ that terminates on $\mathcal{S}_{-1} = \mathcal{F}(\partial\mathcal{M})$. But each such curve $S'$ is in turn part of some curve $\tilde{S}'$ that terminates on $\partial\mathcal{M}$ (we have seen it in the first part of the proof). Therefore, the statement also holds for every curve $S \subset \mathcal{S}_{-n-1} \setminus \mathcal{S}_0$. $\square$

**Corollary 4.4.10.** *For each $n', n'' \geq 0$ the set $\mathcal{M} \smallsetminus (\mathcal{S}_{n'} \cup \mathcal{S}_{n''})$ is a finite or countable union of open domains with piecewise smooth boundaries (curvilinear polygons) such that the interior angles made by their boundary components do not exceed $\pi$ (i.e. these polygons are 'convex' as far as the interior angles are concerned). Some interior angles may be equal to zero.*

*Proof.* By Proposition 4.4.9, the set $\mathcal{S}_{-n'} \cup \mathcal{S}_{n''}$ constitutes a sort of 'net' whose wires end on $\partial\mathcal{M}$. Furthermore, the curves in $\mathcal{S}_{-n'}$ have slope inside the narrow unstable cones $\hat{\mathcal{C}}^u_x$ and those in $\mathcal{S}_{n''}$ have slope in the narrow stable cones $\hat{\mathcal{C}}^s_x$; therefore their intersection forms angles lower than $\pi$. □

The next figure shows a possible realization of $\mathcal{S}_{-2} \cup \mathcal{S}_2$. One can verify that the property of the continuation of the singularity lines holds, for example. Also, the first curvilinear polygons appear. In figuring the set $\mathcal{S}_\infty \cup \mathcal{S}_{-\infty}$ one has to imagine at a version of the next figure that fills densely the domain.

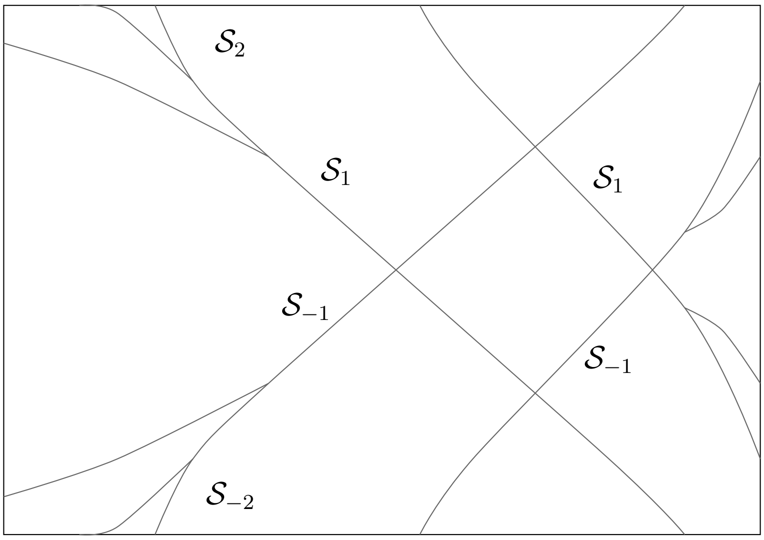


Figure 9: the singularity set $\mathcal{S}_2 \cup \mathcal{S}_{-2}$.

### 4.4.3 Stable and unstable manifolds

Consider the sets

$$\mathcal{S}_\infty = \bigcup_{n=0}^{\infty} \mathcal{S}_n \quad \textit{and} \quad \mathcal{S}_{-\infty} = \bigcup_{n=0}^{\infty} \mathcal{S}_{-n} \tag{4.4.4}$$

of points where *some future* and, respectively, *some past* iterate of $\mathcal{F}$ is singular. We now define stable and unstable manifolds as in [CM]. Often, (un)stable manifolds are defined as curves whose tangent space is everywhere equal to the (un)stable space and which are sets of point having the property that the distance between (past) future trajectories goes to zero for any couple of points. Here we present an alternative construction in which singularity curves play a key role. Our approach has the advantage that the sets of (un)stable manifolds will be by definition a (measurable!) partition of the space $\mathcal{M}$ and that it allows us to directly estimate the size of unstable manifolds (see the next section)[9]. Remember that the whole idea is to find curves that

a) are tangent everywhere to the (un)stable spaces;

b) are smooth connected curves which are never crossed by any singularity line;

c) are maximal.

One could think heuristically in the following way: take a 'good point' $x$ in which the unstable, say, direction $E^u_x$ is defined. Then draw an infinitesimal curve tangent to it, and elongate it in a way that it stays tangent to the unstable spaces of the points it pass trough; then continue elongating until it meet $\mathcal{S}_{-\infty}$. Now a question. Is this always possible? Clearly it is not, since we may choose directly $x \in \mathcal{S}_{-\infty}$. But it could also be that $x$ is in the closure of the set $\mathcal{S}_{-\infty}$ and we are not able to draw any curve starting at $x$ having the right direction, because any such curve would results in an intersection with $\mathcal{S}_{-\infty}$. The very bad news is that actually the closure of the set $\mathcal{S}_{-\infty}$ is all $\mathcal{M}$.

**Lemma 4.4.11.** *Both sets $\mathcal{S}_{\infty}$ and $\mathcal{S}_{-\infty}$ are dense in $\mathcal{M}$.*

*Proof.* Assume, for example that $\mathcal{S}_{\infty}$ is not dense. Then an unstable curve $W \subset \mathcal{M} \setminus \mathcal{S}_{\infty}$ should exists and its images $\mathcal{F}^n(W)$ would grow exponentially fast without ever running into singularities and braking up, so sooner or later they would have to extend beyond the natural boundaries of $\mathcal{M}$. This is due to the fact that their slope lies into the smaller unstable cones (see Proposition 4.4.4) and so it is bounded from above and below, i.e.

$$0 < \mathcal{V}_{min} < d\phi/dr < \mathcal{V}_{max} < \infty. \tag{4.4.5}$$

Therefore, we have a contradiction. □

On the other hand, being in the closure of $\mathcal{S}_{-\infty}$ for a point $x$ does not exclude the possibility of being able to find a curve that pass in $x$, has positive length and does not intersect $\mathcal{S}_{-\infty}$. Indeed consider the following *toy model.*

**Example 1.** Suppose $\mathcal{M} = [0,1]^2$ and $\mathcal{S}_{-\infty} = \Big\{\{x\} \times [0,1] | x \in [0,1] \cap \mathbb{Q}\Big\}$, i.e. the singularity set $\mathcal{S}_{-\infty}$ is made by vertical lines that intersect the up and down side of the square in rational points. Clearly $\mathcal{S}_{-\infty}$ is dense as Lemma 4.4.11 requires. Suppose that

[9] But it has the drawback of giving LSUMs which are not homogeneous in the sense that their distortion under iterations of $\mathcal{F}$ is not bounded, see Chapter 6.

the unstable direction $E^u_x$ is everywhere the vertical one (in real examples $E^u_x$ varies with the point usually). Now if the game is drawing a line (which represent the LUM) in the vertical direction until it meets some singularity line, then we can do this for every point $(x, y) \in [0, 1]^2$ which has $x \notin \mathbb{Q} \cap [0, 1]$. Therefore, even if $\mathcal{S}_{-\infty}$ is dense, almost every point in $\mathcal{M}$ has an 'unstable manifold' passing trough him, and the length of such unstable manifolds (vertical lines) is almost everywhere equal to one!

Although the previous example is a big simplification (there, the singularity curves never *cut* anything), the reality is not so much different. The crucial property is that singularity curves tends to align to the (un)stable direction (see Proposition 4.4.7 for the alignment property of singularity curves), so that the real case become surprisingly similar to the silly Example (1) where singularity curves do not intersect the LSUMs because they are parallel to them. In the reality this starts to happen for large $n$ for the sets $\mathcal{S}_{-n} \setminus \mathcal{S}_{-(n-1)}$; furthermore the *cutting* actually never stops but it interests as $n \to \infty$ a smaller and smaller measure of points. Therefore, we can really say that it is the alignment property the physical mechanism that will ultimately lead to the existence almost everywhere of the LSUMs [10].
Alignment property has an additional advantage, actually it implies that we can define the LSUMs only trough property (b) and (c), while (a) will be a consequence. Indeed, heuristically, if we require that our, say, LUM $W^u$, never cross the singularity curves $\mathcal{S}_{-\infty} = \bigcup_{n\geq 0} \mathcal{S}_{-n}$ which are dense, the only way we can work it out is by making $W^u$ as parallel as possible to the tail of $\mathcal{S}_{-\infty}$, i.e. $\bigcup_{n\geq N} \mathcal{S}_{-n}$ for bigger and bigger $N$. But we know that this tail will start to align to the unstable spaces (thanks to alignment, again). Therefore, also $W^u$ will, since it has to be 'parallel'[11]. Let us now define exactly LSUMs and we will address more in detail the problem of their existence later on.
Observe that $\mathcal{S}_\infty$ is a countable union of smooth stable curves and the boundary $\partial\mathcal{M}$, while $\mathcal{S}_{-\infty}$ is a countable union of smooth unstable curves, plus $\partial\mathcal{M}$. For each $\mathcal{M}_i \subset \mathcal{M}$ (cf section 4.2.2), the sets $\mathcal{M}_i \cap \mathcal{S}_\infty$ and $\mathcal{M}_i \cap \mathcal{S}_{-\infty}$ are pathwise connected, by Proposition 4.4.9. On the other hand, the set $\mathcal{M} \setminus \mathcal{S}_{-\infty}$ has full $\mu$ measure, but it is badly disconnected. We analyze its connected components next. Let $x \in \mathcal{M} \setminus \mathcal{S}_{-\infty}$ and for $n \geq 1$ denote by $Q_{-n}(x)$ the connected component of the open set of $\mathcal{M} \setminus \mathcal{S}_{-n}$ that contains $x$. By Corollary 4.4.10, one has that $Q_{-n}(x)$ is a curvilinear polygon with interior angles $\leq \pi$. Obviously, $Q_{-n}(x) \supseteq Q_{-(n+1)}(x)$ for all $n \geq 1$ and the intersection of their closures

$$\tilde{W}^u(x) := \bigcap_{n=1}^{\infty} \overline{Q}_{-n}(x)$$

is a closed continuous monotonically increasing curve.

[10] The mathematical jargon says that *alignment wins over singularities.*

[11] Certainly parallel is not mathematically correct. The mathematically correct statement is precisely the one of Proposition 4.4.7.

**Definition 4.4.12.** The curve $\tilde{W}^u(x)$ without its endpoints is called *(local) unstable manifold* at $x$ and we denote it by $W^u(x)$.

**Lemma 4.4.13.** *We have* $W^u(x) \subseteq \bigcap_{n\geq 0} Q_{-n}(x)$.

*Proof.* We have that $\partial Q_{-n}(x)$ consists of two monotonically increasing and piecewise smooth curves whose endpoints are the top and bottom vertices of $Q_{-n}(x)$. We call those curves left and right sides and denote them by $\partial^L Q_{-n}(x)$ and $\partial^R Q_{-n}(x)$, respectively. Suppose that some point $y \in W^u(x)$ belongs to a singularity curve $S \subset \partial Q_{-n}(x)$ for a finite $n \geq 1$. Without loss of generality, we assume $S \subset \partial^L Q_{-n}(x)$. Then $y$ is also a limit point for a sequence of curves $S'_{-m} \subset \partial^R Q_{-m}(x)$. Since $S \subset \mathcal{S}_{-n}$ for a finite $n$, the curve $S$ has a slope at $y$ different from the limit slope of the curves $S'_{-m}$. Indeed, it can be proved that (as we saw in (4.4.7)) the singularity curves of $\mathcal{S}_{-n} \setminus \mathcal{S}_{-(n-1)}$ have tangent space that align everywhere to the unstable direction as $n \to \infty$ but it is different from it for any finite $n$. This leads to a contradiction. $\square$

**Proposition 4.4.14.** *At every point* $y \in W^u(x)$ *the slope of the curve* $W^u(x)$ *equals* $\mathcal{V}^u(y)$.

*Proof.* For every $y \in W^u(x)$ we can find curves $S_{-m} \subset \mathcal{S}_{-m}$ and $y_m \in S_{-m}$ such that $y_{-m} \to y$ as $m \to \infty$. By alignment one has that $\mathcal{V}(y_m) \to \mathcal{V}^u(y)$ as $m \to \infty$. $\square$

One can also prove the following theorem on the regularity of the curves $W^u(x)$, whose proof we omit.

**Theorem 4.4.15.** *The curve* $W^u(x)$ *is* $C^{l-2}$ *smooth, all its* $l-2$ *derivatives are bounded by a constant independent of* $x$*, and its* $(l-2)$*nd derivative is Lipschitz continuous with a Lipschitz constant independent of* $x$.

Observe that $\mathcal{F}^{-n}(W^u(x)) \subset \mathcal{M} \setminus \mathcal{S}_{-\infty}$ is an unstable curve for every $n \geq 1$. The uniform hyperbolicity of $\mathcal{F}$ (see section (4.4.1)) implies

$$|\mathcal{F}^{-n}(W^u(x))| \leq C\Lambda^{-n}|W^u(x)|, \quad \forall n \geq 1,$$

for some $\Lambda > 1$ hence the preimages of $W^u(x)$ contracts exponentially fast.
All our constructions have their time-reversal. In particular, for every point $x \in \mathcal{M} \setminus \mathcal{S}_\infty$ we have a stable curve $W^s(x)$ which satisfies

$$|\mathcal{F}^{n}(W^s(x))| \leq C\Lambda^{-n}|W^s(x)| \quad \forall n \geq 1.$$

This last fact remind us the *usual* definition of stable and unstable manifolds.

**Definition 4.4.16.** (Alternative definition of LSUM) A smooth curve $W^u \subset \mathcal{M}$ is called an *unstable manifold* if the map $\mathcal{F}^{-n}$ is smooth on $W^u$ for every $n \geq 1$ and

$$\lim_{n\to\infty} |\mathcal{F}^{-n}(W^u)| = 0.$$

Similarly, a smooth curve $W^s \subset \mathcal{M}$ is called a *stable manifold* if the map $\mathcal{F}^n$ is smooth on $W^s$ for every $n \geq 1$ and

$$\lim_{n\to\infty} |\mathcal{F}^n(W^s)| = 0.$$

In our case, our analysis shows that $W^u(x)$ and $W^s(x)$ are the maximal unstable and, respectively, stable manifolds passing trough $x$ (a property that some authors include in the definition), and the convergence in the above definition is actually exponentially fast.

### 4.4.4 Size of unstable manifolds

We now want to show that $\tilde{W}^u(x) := \bigcap_{n=1}^{\infty} \overline{Q}_{-n}(x)$ is a curve with a positive diameter passing trough $x$ for $\mu$-almost every $x \in \mathcal{M}$. The powerful method that allow us to state that is the *Borel-Cantelli Lemma*.

**Lemma 4.4.17.** *(Borel-Cantelli). Let $(X, \mathfrak{B}, \mu)$ be a measure space and $\{B_n\}_{n\in\mathbb{N}}$ a sequence of measurable events. Suppose that*

$$\sum_{n=1}^{\infty} \mu(B_n) < \infty$$

*Then, the measurable set $B := \bigcap_{m\geq 1} \bigcup_{n\geq m} B_n$ has zero measure.*

Let $W \subset \mathcal{M}$ be a smooth unstable (or stable) curve. Any point $x \in W$ divides $W$ into two segments, and we denote by $r_W(x)$ the length (in the Euclidean metric) of the shorter one. For brevity, we put $r^u(x) := r_{W^u(x)}(x)$. If it happens that $W^u(x) = \emptyset$, then we set $r^u(x) = 0$. Clearly, $r^u(x)$ characterizes the size of $W^u(x)$.

**Theorem 4.4.18.** *$W^u(x)$ exists (i.e. $r^u(x) > 0$) for almost every $x \in \mathcal{M}$. Furthermore,*

$$\mu(\{x | r^u(x) < \epsilon\}) \leq C\epsilon, \tag{4.4.6}$$

*for some constant $C = C(\mathcal{D}) > 0$ and all $\epsilon > 0$.*

**Remark 4.4.19.** Actually, the existence of $W^u(x)$ follows from the general theory of hyperbolic maps with singularities (see [KS]) but the estimate (4.4.6) is billiard specific. The general theory guarantees only that $\mu(\{x|r^u(x) < \epsilon\}) \leq C\epsilon^a$ for some $a > 0$.

*Proof.* For any point $x \in \mathcal{M}$ denote by $d^u(x, \mathcal{S}_1)$ the length of the shortest unstable curve that connects $x$ with the set $\mathcal{S}_1$. We begin proving the following lemma.

**Lemma 4.4.20.** *For any $x \in \mathcal{M}$*

$$r^u(x) \geq \min_{n\geq 1} \hat{c}\Lambda^n d^u(\mathcal{F}^{-n}x, \mathcal{S}_1), \tag{4.4.7}$$

*where $\hat{c} = \hat{c}(\mathcal{D})$ is a constant from (4.4.2).*

*Proof.* We may assume that $x \in \mathcal{M} \setminus \mathcal{S}_{-\infty}$. Denote $x_{-n} = \mathcal{F}^{-n}(x)$ for $n \geq 1$ and again let $Q_{-n}(x)$ be the connected component of $\mathcal{M} \setminus \mathcal{S}_{-n}$ containing the point $x$. Obviously, $Q_n(x_{-n}) := \mathcal{F}^{-n}(Q_{-n}(x))$ is the connected component of $\mathcal{M} \setminus \mathcal{S}_n$ containing the point $x_{-n}$. Let $W'_{-n}$ be an arbitrary unstable curve passing trough $x_{-n}$ and terminating on the opposite sides of $Q_n(x_{-n})$. Then $W' = \mathcal{F}^n(W'_{-n})$ is an unstable curve passing trough $x$ and terminating on $\partial Q_{-n}(x)$. It is divided by the point $x$ into two segments, and we denote by $W$ the shorter one. Put $W_{-m} = \mathcal{F}^{-m}(W)$ for all $m \geq 1$. Since $W_{-n}$ terminates on $\mathcal{S}_n$, there is an $m \in [1, n]$ such that $W_{-m}$ joins $x_{-m}$ with $\mathcal{S}_1$. Due to the uniform hyperbolicity of $\mathcal{F}$, we have

$$r_{W'}(x) = |W| \geq \hat{c}\Lambda^m |W_{-m}| \geq \hat{c}\Lambda^m d^u(\mathcal{F}^{-m}x, \mathcal{S}_1).$$

Since this holds for all $n \geq 1$, the lemma follows. □

Next, we pick $\hat{\Lambda} \in (1, \Lambda)$ and observe that

$$r^u(x) \geq r^u_*(x) := \min_{n \geq 1} \hat{c}\hat{\Lambda}^n d^u(\mathcal{F}^{-n}, \mathcal{S}_1),$$

(we need to assume $\hat{\Lambda} < \Lambda$ for a later use). So if $r^u(x) = 0$, then $r^u_*(x) = 0$; i.e. the past semitrajectory of $x$ approaches the singularity set $\mathcal{S}_1$ faster than the exponential function $\hat{\Lambda}^{-n}$. We employ the Borel-Cantelli lemma to show that this happens with probability zero. For any $\epsilon > 0$ denote by $\mathcal{U}_\epsilon(\mathcal{S}_1)$ the $\epsilon$-neighborhood of the set $\mathcal{S}_1$ and let

$$\mathcal{U}^u_\epsilon(\mathcal{S}_1) = \{x | d^u(x, \mathcal{S}_1) < \epsilon\},$$

this is a sort of $\epsilon$-neighborhood of $\mathcal{S}_1$, measured along unstable curves. Since the tangent to the unstable curves lie in the unstable cones (see (4.4.5)), one has

$$\mathcal{U}^u_\epsilon(\mathcal{S}_1) \subset \mathcal{U}_{D\epsilon}(\mathcal{S}_1),$$

for some constant $D > 0$ (because the set $\mathcal{S}_1$ consists of monotonically decreasing curves and horizontal lines); hence

$$\mu(\mathcal{U}^u_\epsilon(\mathcal{S}_1)) < C\epsilon, \tag{4.4.8}$$

for some $C > 0$ and all $\epsilon > 0$. Next we show that $r^u_*(x) > 0$ for $\mu$-almost every point $x \in \mathcal{M}$. Let

$$B_n := \mathcal{F}^n(\mathcal{U}^u_{\hat{\Lambda}^{-n}}(\mathcal{S}_1)) = \{x | d^u(\mathcal{F}^{-n}x, \mathcal{S}_1) \leq \hat{\Lambda}^{-n}\}.$$

By the invariance of the measure $\mu$ we have

$$\sum_{n=1}^{\infty} \mu(B_n) = \sum_{n=1}^{\infty} \mu(\mathcal{U}^u_{\hat{\Lambda}^{-n}}(\mathcal{S}_1)) < \infty.$$

Denote by $B = \bigcap_{m \geq 1} \bigcup_{n \geq m} B_n$ the event 'infinitely-often'-$B_n$. The Borel-Cantelli lemma implies that $\mu(B) = 0$, and it is easy to check that $r^u_*(x) > 0$ for every $x \in \mathcal{M} \setminus B$.

**Lemma 4.4.21.** *We have* $\mu(\{x|r_*^u(x) < \epsilon\}) \leq C\epsilon$ *for some constant* $C = C(\mathcal{D})$ *and all* $\epsilon > 0$.

*Proof.* Observe that $r_*^u(x) < \epsilon$ iff $x \in \mathcal{F}^n(\mathcal{U}^u_{\epsilon\hat{c}^{-1}\hat{\Lambda}^{-n}}(\mathcal{S}_1))$ for some $n \geq 1$. Now (4.4.8) implies the result. □

This complete the proof of Theorem 4.4.18. □

Yet another way to understand the previous proof is considering the following facts.

a) one can show that almost every point approaches the singularity set $\mathcal{S}_1$ at most polinomially, i.e. for almost every $x$ there exists a constant $H = H(x) > 0$ such that

$$d^u(\mathcal{F}^{-i}x, \mathcal{S}_1) \geq \frac{H(x)}{(1+i)^e},$$

for some $e > 0$.

b) the fact that LUMs are curves that are unstable at all orders (i.e., $\mathcal{F}^{-n}(W)$ is unstable for all $n \geq 0$ if $W$ is a LUM) and the fact that along the backward trajectory of $x$ any unstable curve contracts exponentially, imply that there can only be a finite number of cuts (indeed, the polynomial speed may win over the exponential contraction only a finite number of times).

To get part (a), one uses Borel-Cantelli lemma and the fact that the measure of the $\epsilon$-neighborhood of singularity lines does not exceed $C\epsilon^a$ for some positive $C$ and $a$. For billiards $a = 1$.

# Chapter 5

# A Direct Proof of the K-Property for the Baker's Map

In order to illustrate the core ideas for the proofs of section 6, which are rather technical, here we give a simple proof of the fact that the Baker's Map is a K-automorphism.

## 5.1 Kolmogorov automorphisms

**Definition 5.1.1.** An invertible measure-preserving transformation $T$ of a probability space $(X, \mathfrak{F}, m)$ is a *Kolmogorov automorphism* (K-automorphism) if there exists a sub-$\sigma$-algebra $\mathfrak{B}$ of $\mathfrak{F}$ such that:

a) $\mathfrak{B} \subseteq T\mathfrak{B}$ .

b) $\bigvee_{n=0}^{\infty} T^n \mathfrak{B} = \mathfrak{F}$.

c) $\bigcap_{n=0}^{\infty} T^{-n} \mathfrak{B} = \mathfrak{N} = \{A | m(A) = 0 \text{ or } m(X \setminus A) = 0\}$.

We always assume $\mathfrak{N} \neq \mathfrak{F}$ (since if not the identity is the only measure-algebra automorphism). Hence $\mathfrak{B} \neq T\mathfrak{B}$.

## 5.2 Definition of the map

Let $(\mathbb{T}^2, Borel(\mathbb{T}^2), Leb_2)$ be the probability space on the two dimensional torus equipped with the Borel $\sigma$-algebra and the two dimensional Lebesgue measure. Let $(x, y)$ be a point in $\mathbb{T}^2$, we define the *Baker's map* as

$$T(x, y) = \begin{cases} (2x, y/2), & \text{if } 0 \leq x < 1/2 \\ (2 - 2x, 1 - y/2), & \text{if } 1/2 \leq x < 1. \end{cases} \tag{5.2.1}$$

One can show that $T$ preserves $Leb_2$.

## 5.3 Some basic facts about $\sigma$-algebras and partitions

We are now going to review the most important notions on $\sigma$-algebras and partitions. These will play a fundamental role in the development of the results of the report. Sometimes it is convenient to define a $\sigma$-algebra specifying only some of the sets that it includes. The minimum $\sigma$-algebra that includes the system of sets $\mathcal{G}$ is denoted by $\sigma(\mathcal{G})$ and it is defined as

$$\sigma(\mathcal{G}) = \bigcap_{\substack{\mathfrak{F}\ \sigma\text{-algebra} \\ \mathfrak{F} \supseteq \mathcal{G}}} \mathfrak{F}. \tag{5.3.1}$$

The proof of (5.3.1) rests on the basic fact that an intersection of $\sigma$-algebras is a $\sigma$-algebra itself. Let $(X, \mathfrak{F}, \mu)$ and $(Y, \mathfrak{G}, \nu)$ be two $\sigma$-finite measure spaces. A problem in defining a product measure space is that the collection $\mathfrak{F} \times \mathfrak{G}$ is in general not a $\sigma$-algebra but only a semi-ring.

**Definition 5.3.1.** We define the *product $\sigma$-algebra* of two $\sigma$-algebras $\mathfrak{F}$ and $\mathfrak{G}$ as the coarsest $\sigma$-algebra that contains $\mathfrak{F} \times \mathfrak{G}$ and we denote it by $\mathfrak{F} \otimes \mathfrak{G}$.

$$\mathfrak{F} \otimes \mathfrak{G} := \sigma(\mathfrak{F} \times \mathfrak{G}). \tag{5.3.2}$$

A standard theorem, whose we omit the proof (see e.g. [Sc] chapter 13), states the existence and uniqueness of product measures.

**Theorem 5.3.2.** *Let $(X, \mathfrak{F}, \mu)$ and $(Y, \mathfrak{G}, \nu)$ be two $\sigma$-finite measure spaces. Then the set function*

$$\rho : \mathfrak{F} \times \mathfrak{G} \to [0, \infty], \quad \rho(A, B) := \mu(A)\nu(B), \tag{5.3.3}$$

*extends uniquely to a measure on $\mathfrak{F} \otimes \mathfrak{G}$.*

**Remark 5.3.3.** The $\sigma$-finiteness assumption in the theorem above is necessary to ensure the uniqueness of the measure $\rho$.

**Definition 5.3.4.** Let $(X, \mathfrak{F}, \mu)$ and $(Y, \mathfrak{G}, \nu)$ be two $\sigma$-finite measure spaces. The unique measure $\rho$ constructed in Theorem 5.3.2 is called the *product* of the measures $\mu$ and $\nu$ and is denoted by $\mu \times \nu$. $(X \times Y, \mathfrak{F} \otimes \mathfrak{G}, \mu \times \nu)$ is called the *product measure space.*

**Corollary 5.3.5.** *(of Theorem 5.3.2) If $n > d \geq 1$, then*

$$(\mathbb{R}^n, Borel(\mathbb{R}^n), Leb_n) = (\mathbb{R}^d \times \mathbb{R}^{n-d}, Borel(\mathbb{R}^d) \otimes Borel(\mathbb{R}^{n-d}), Leb_d \times Leb_{n-d}). \tag{5.3.4}$$

Let $(X, \mathfrak{F})$ be a measurable space. We will always implicitly assume that the elements of a partitions are measurable sets. $\xi(x)$ is the element of $\xi$ containing $x$. A set $A \in \mathfrak{F}$ is called $\xi$-saturated if $x \in A$ implies $\xi(x) \subseteq A$. Thus $A = \bigcup_{x \in A} \xi(x)$. A partition $\xi$ induces a sub-$\sigma$-algebra of measurable $\xi$-saturated sets,

$$\mathfrak{F}(\xi) = \{A \in \mathfrak{F} : \text{ A is } \xi\text{-saturated}\}. \tag{5.3.5}$$

If by $\xi$ we mean a set of equal mod 0 partitions, also $\mathfrak{F}(\xi)$ will stand for the correspondent equivalence class of mod 0 equal sub-$\sigma$-algebras. From now on, we will adopt this point of view. Furthermore, without loss of generality, we will consider only (classes of equal mod 0) *complete* sub-$\sigma$-algebras.

Let $(X, \mathfrak{F})$ be a measurable space and suppose we have a sequence of sub-$\sigma$-algebras each one correspondent to a partition, i.e. $\mathfrak{F}_1 = \mathfrak{F}(\xi_1)$, $\mathfrak{F}_2 = \mathfrak{F}(\xi_2)$, ... .We define their *join* as the coarsest sub-$\sigma$-algebra that contains all the $\mathfrak{F}_m$, and we denote it by $\bigvee_m \mathfrak{F}_m$. In particular

$$\bigvee_m \mathfrak{F}_m = \bigcap_{\substack{\mathfrak{G} \text{ sub-}\sigma\text{-algebra of } \mathfrak{F} \\ \mathfrak{G} \supseteq \mathfrak{F}_m,\ \forall m}} \mathfrak{G}. \tag{5.3.6}$$

Symmetrically, the finest sub-$\sigma$-algebra that is coarser than each $\mathfrak{F}_m$ is simply the intersection $\bigcap_m \mathfrak{F}_m$. Thanks to (5.3.5), we associate to these *maximal* and *minimal* $\sigma$-algebras two partitions which we will denote $\bigvee_m \xi_m$ and $\bigwedge_m \xi_m$ (the *join partition* and *meet partition*), defined as

$$\mathfrak{F}(\bigvee_m \xi_m) = \bigvee_m \mathfrak{F}(\xi_m), \tag{5.3.7}$$

$$\mathfrak{F}(\bigwedge_m \xi_m) = \bigwedge_m \mathfrak{F}(\xi_m). \tag{5.3.8}$$

**Remark 5.3.6.** The *join* operation (both on sub-$\sigma$-algebras and partitions) is defined only with respect to some *ambient* $\sigma$-algebra. Indeed, the result of eq. (5.3.6) may change, changing $\mathfrak{F}$. In particular, $\bigvee_m \mathfrak{F}_m \subseteq \mathfrak{F}$.

## 5.4 The Proof

Let $\xi^u$ and $\xi^s$ be the partitions of $\mathbb{T}^2$, whose elements are the horizontal and vertical segments on the torus, respectively.

$$\xi^s = \Big\{ \{(x,y) | y \in [0,1)\}_{x \in [0,1)} \Big\}, \tag{5.4.1}$$

$$\xi^u = \Big\{ \{(x,y) | x \in [0,1)\}_{y \in [0,1)} \Big\}. \tag{5.4.2}$$

For reasons that will be discussed later, we may call the elements of $\xi^u$ unstable manifolds and those of $\xi^s$ stable manifolds. The correspondent $\sigma$-algebras will be denoted $\mathfrak{B}_u$ and $\mathfrak{B}_s$. Call $\pi_x$ and $\pi_y$ the projections on the horizontal and vertical sides of the torus, respectively. By (5.3.4), we can decompose our measure space in the following way

$$(\mathbb{T}^2, Borel(\mathbb{T}^2), Leb_2) = (\mathbb{T} \times \mathbb{T}, Borel(\mathbb{T}) \otimes Borel(\mathbb{T}), Leb_1 \times Leb_1). \tag{5.4.3}$$

By the correspondence between $\sigma$-algebras and partitions, we may define (mod 0) the following sub-$\sigma$-algebras of $Borel(\mathbb{T}^2)$:

$$\mathfrak{B}_{\mathfrak{s}} = \{A \in Borel(\mathbb{T}^2) |\ \pi_x(A) \in Borel(\mathbb{T}) \text{ and } \pi_y(A) = [0,1)\} = Borel(\mathbb{T}) \otimes \mathfrak{N}_2, \quad (5.4.4)$$

$$\mathfrak{B}_{\mathfrak{u}} = \{A \in Borel(\mathbb{T}^2) |\ \pi_y(A) \in Borel(\mathbb{T}) \text{ and } \pi_x(A) = [0,1)\} = Borel(\mathbb{T}) \otimes \mathfrak{N}_1, \quad (5.4.5)$$

where $\mathfrak{N}_{1(2)}$ is the trivial $\sigma$-algebra on $[0,1)$ on the $x(y)$ axis.

Now, let us consider the partitions of the unit interval $\xi_m$

$$\xi_m = \left\{ \Big[0, \frac{1}{2^m}\Big), \Big[\frac{1}{2^m}, \frac{2}{2^m}\Big), \ \dots\ , \ \Big[\frac{2^m-1}{2^m}, 1\Big) \right\}, \quad (5.4.6)$$

and the correspondent sub-$\sigma$-algebra $\mathfrak{B}(\xi_m)$, with respect to $Borel([0,1))$. For example $\xi_1 = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ and $\mathfrak{B}(\xi_1) = \{\emptyset, [0,1), [0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. These observations will be useful later on.

**Lemma 5.4.1.** $\bigvee_m \mathfrak{B}(\xi_m) = Borel([0,1))$.

*Proof.* We only prove that $\bigvee_m \mathfrak{B}(\xi_m) \supseteq Borel([0,1))$, since the other inclusion was discussed in Remark 5.3.6. Let $\mathcal{I}_{dyadic}$ be the collection of open sets of the form $(a,b)$ where $a$ and $b$ are dyadic numbers, i.e. $a = \frac{c}{2^m}$ and $b = \frac{d}{2^n}$, $a < b \in [0,1)$, $c, d, m, n \in \mathbb{N}^+$. Clearly $\bigvee_m \mathfrak{B}(\xi_m) \supseteq \sigma(\mathcal{I}_{dyadic})$. Indeed, let $[a,b)$ be a set with $a$ and $b$ dyadic as before. There exists an $m \in \mathbb{N}^+$ sufficiently big such that $[a,b) \in \mathfrak{B}(\xi_m)$. But

$$(a,b) = \bigcup_{j \in \mathbb{N}} [a + 1/2^j, b), \quad (5.4.7)$$

and

$$[a,b) = \bigcap_{j \in \mathbb{N}} (a - 1/2^j, b). \quad (5.4.8)$$

These formulae imply that the two $\sigma$-algebras generated by open and half-open sets actually coincide.

By definition $Borel([0,1)) = \sigma(O)$, where $O$ is the collection of open subset of $[0,1)$. Clearly $\sigma(O) \supseteq \sigma(\mathcal{I}_{dyadic})$. On the other hand if $U \in O$ then

$$U = \bigcup_{\substack{I \in \mathcal{I}_{dyadic} \\ I \subseteq U}} I. \quad (5.4.9)$$

Here '$\supseteq$' is clear from the definition of $U$. For the other direction '$\subseteq$' we fix $x \in U$. Since $U$ is open, there is some open ball $B_\epsilon(x) \subseteq U$ and we can shrink it to get an internal set

$I = I(x) \in \mathcal{I}_{dyadic}$ containing $x$. Since every dyadic number is rational there are at most $(\#\mathbb{Q} \times \#\mathbb{Q}) = \#\mathbb{N}$ many $I$ in the union (5.4.9) and

$$U = \bigcup_{x\in U}\{x\} \subseteq \bigcup_{x\in U} I(x) = \bigcup_{\substack{I \text{ dyadic} \\ I\subset U}} I.$$

Thus

$$U \in O \subseteq \sigma(\mathcal{I}_{dyadic}), \tag{5.4.10}$$

proving the other inclusion. Hence $\bigvee_m \mathfrak{B}(\xi_m) \supseteq \sigma(\mathcal{I}_{dyadic}) = Borel([0,1))$ □

**Lemma 5.4.2.** *Let $A \in Borel(\mathbb{T})$. Suppose there exists a sequence of real numbers $\{\delta_n\}_{n\in\mathbb{N}^+}$ such that:*

*a)* $\delta_n \searrow 0$*,*

*b)* $A + \delta_n = A$ *mod 1* $\forall n$*.*

*Then $A$ has either zero or full measure.*

*Proof.* (Sketch) Let $A \subseteq [0,1)$ be a Borel measurable set such that $0 < \mu(A) < 1$. Pick a Lebesgue density point $x \in A$ and a Lebesgue density point $y \in [0,1) \setminus A$ and translate a small neighborhood of $x$ into a small neighborhood of $y$ arriving at a contradiction. □

**Theorem 5.4.3.** *The Baker's map is a Kolmogorov Automorphism.*

*Proof.* We know that $T\xi^s = \{TA|A \in \xi^s\} = \Big\{\{(x,y)|y \in [0,1/2)\}_{x\in[0,1)}, \{(x,y)|y \in (1/2,1)\}_{x\in[0,1)}\Big\}$. On the other hand one can verify that $T^{-1}\xi^s = \Big\{\{(x,y) \cup (x + 1/2 \text{ mod } 1)|y \in [0,1)\}_{x\in[0,1)}\Big\}$. By an inductive argument

$$T^m\xi^s = \Big\{\{(x,y)|y \in [0,1/2^m)\}_{x\in[0,1)}, ..., \{(x,y)|y \in ((2^m-1)/2^m, 1)\}_{x\in[0,1)}\Big\} \tag{5.4.11}$$

and

$$T^{-m}\xi^s = \Big\{\{(x,y) \cup (x + 1/2^m \text{ mod } 1)|y \in [0,1)\}_{x\in[0,1)}\Big\}, \tag{5.4.12}$$

for $m \in \mathbb{N}$. $\forall n \in \mathbb{Z}$, we denote by $T^n\mathfrak{B}_s$ the correspondent (mod 0) $\sigma$-algebras. Now, by (5.4.11), $\forall m \in \mathbb{N}^+$:

$$T^m\mathfrak{B}_{\mathfrak{s}} = Borel(\mathbb{T}) \otimes \mathfrak{B}(\xi_m). \tag{5.4.13}$$

Property (a) of Definition 5.1.1 is easily proven:

$$T\mathfrak{B}_s = Borel(\mathbb{T}) \otimes \mathfrak{B}(\xi_1) \supseteq Borel(\mathbb{T}) \otimes \mathfrak{N}_2 = \mathfrak{B}_s. \tag{5.4.14}$$

Regarding property (b), we have

$$\begin{aligned}&\bigvee_m T^m\mathfrak{B}_s = \bigvee_m \Big(Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m)\Big) = Borel(\mathbb{T})\otimes\bigvee_m \mathfrak{B}(\xi_m) = \\ &Borel(\mathbb{T})\otimes Borel(\mathbb{T}) = Borel(\mathbb{T}^2),\end{aligned} \tag{5.4.15}$$

where in the third equality we used Lemma 5.4.1. Let us stop for a moment to prove the second equality. By definition,

$$\bigvee_m \Big(Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m)\Big) = \bigcap_{\substack{\mathfrak{G}\text{ sub-}\sigma\text{-algebra}\\ \mathfrak{G}\supseteq Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m),\ \forall m}} \mathfrak{G}. \tag{5.4.16}$$

Note that, in the intersection above, for each $\mathfrak{G}$, $\mathfrak{G} \supseteq Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m) \supseteq Borel(\mathbb{T})\times\mathfrak{B}(\xi_m)$ $\forall m$. Hence $\mathfrak{G} \supseteq Borel(\mathbb{T})\times\bigvee_m\mathfrak{B}(\xi_m)$ because $\mathfrak{G}$ is a $\sigma$-algebra. It follows that $\mathfrak{G} \supseteq Borel(\mathbb{T})\otimes\bigvee_m\mathfrak{B}(\xi_m)$, again because $\mathfrak{G}$ is a $\sigma$-algebra. This implies that $\bigvee_m Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m) \supseteq Borel(\mathbb{T})\otimes\bigvee_m\mathfrak{B}(\xi_m)$. On the other hand,

$$Borel(\mathbb{T})\otimes\bigvee_m\mathfrak{B}(\xi_m) = \bigcap_{\substack{\mathfrak{F}\text{ is a sub-}\sigma\text{-algebra}\\ \mathfrak{F}\supseteq Borel(\mathbb{T})\times\bigvee_m\mathfrak{B}(\xi_m)}} \mathfrak{F}. \tag{5.4.17}$$

In the intersection above, $\mathfrak{F} \supseteq Borel(\mathbb{T})\times\bigvee_m\mathfrak{B}(\xi_m) \supseteq Borel(\mathbb{T})\times\mathfrak{B}(\xi_m)$ for all $m$. Hence $\mathfrak{F} \supseteq Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m)$ $\forall m$. Therefore, $\mathfrak{F} \supseteq \bigvee_m Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m)$. This means that $\bigvee_m Borel(\mathbb{T})\otimes\mathfrak{B}(\xi_m) \subseteq Borel(\mathbb{T})\otimes\bigvee_m\mathfrak{B}(\xi_m)$. And equality is proved.

Consider now the partitions $\xi_{-m} = T^{-m}\xi$ and the correspondent $\sigma$-algebras $T^{-m}\mathfrak{B}_s$. If $A\in T^{-m}\xi$, by (5.4.12) $A$ is invariant with respect to a shift of $1/2^m$ in the $x$-direction, i.e.

$$\pi_x(A) + \frac{1}{2^m} = \pi_x(A)\ (mod\ 1). \tag{5.4.18}$$

Hence, every set belonging to $\bigcap_{n=0}^{\infty} T^{-n}\mathfrak{B}_s$ must enjoy that symmetry $\forall m$. Indeed, by definition, every such set is $T^{-m}\xi$-saturated for all $m$ (see 5.3.5). Let us take $B \in \bigcap_{n=0}^{\infty} T^{-n}\mathfrak{B}_s$ such that $Leb_2(B) > 0$. Clearly $Leb_1(\pi_x(B)) > 0$. If we choose $\delta_n = \frac{1}{2^n}$, by Lemma 5.4.2, we have that $Leb_1(\pi_x(B)) = 1$ and, since $B\in\mathfrak{B}_s$, we have that $Leb_1(\pi_y(B)) = 1$. By (5.4.3), this imply that $Leb_2(B) = 1$, completing the proof. $\square$

# Chapter 6

# A Structure Theorem for Infinite Measure Maps with Singularities

In this chapter we present the original part of this thesis: a Structure Theorem for infinite measure maps with singularity, which proves the atomicity of the tail $\sigma$-algebra for hyperbolic maps that satisfy assumptions (H1)-(H6) listed below.

## 6.1 Exactness, K-mixing and partitions

The statistical property trough which we get a characterization of our dynamical system is the *K-property* (see Definition 5.1.1 in the previous chapter). Let us spend a few comments about Definition 5.1.1. In any Lebesgue space there is a one to one correspondence between measurable partitions and sub-$\sigma$-algebras. Assuming that $(X, \mathfrak{F}, m)$ is indeed a Lebesgue space, we can use this to re-write entirely Definition 5.1.1 in terms of partitions of the space $X$. Doing so, we have the following definition of *K-mixing*.

**Definition 6.1.1.** A measure-preserving automorphism $T : X \to X$ is said to be *Kolmogorov* (or has the *K-property*, is *K-mixing*) if there exists a measurable partition $\xi$ of $X$ with three properties:

a) $\xi \leq T\xi$;

b) $\bigvee_{n=0}^{\infty} T^n \xi = \epsilon$;

c) $\bigwedge_{n=0}^{\infty} T^{-n} \xi = \nu$,

where $\epsilon$ and $\nu$ are, respectively, the partition into points and the trivial partition. We also remind that when we talk about any partition we actually mean the whole equivalence class of (mod 0) equal partitions.

The second comment is about the relation between *K-mixing* and *exactness*, see e.g. [L2].

**Definition 6.1.2.** The measure-preserving dynamical system $(X, \mathfrak{F}, m, T)$ is called exact if

$$\bigcap_{n=0}^{\infty} T^{-n}\mathfrak{F} = \mathfrak{N} \mod m. \tag{6.1.1}$$

Or, using the language of partitions:

**Definition 6.1.3.** The measure-preserving dynamical system $(X, \mathfrak{F}, m, T)$ is called exact if

$$\bigwedge_{n=0}^{\infty} T^{-n}\epsilon = \nu. \tag{6.1.2}$$

Since exactness implies that $T^{-1}\mathfrak{F} \neq \mathfrak{F}$ mod m, a non-trivial exact $T$ cannot be an automorphism of the measure space $(X, \mathfrak{F}, m)$. On the other hand we can think of K-mixing as being 'morally' the counterpart of exactness for invertible transformation. Indeed, consider the action on some partition $\xi$ induced by $T$. We call such a map $T_\xi : \xi \to \xi$ and we define it by

$$T_\xi \xi(x) = \xi(Tx),$$

where $\xi(z)$, $z \in X$ is the element of $\xi$ that contains $z$. Since we are in a Lebesgue space, there is a $\sigma$-algebra $\mathfrak{F}_\xi$ associated to the measurable partition $\xi$ . Now, suppose that the dynamical system $(X, \mathfrak{F}, m, T)$ is K-mixing with respect to the $\sigma$-algebra $\mathfrak{F}_\xi$, then by property c) of Definition 6.1.1 we have that the dynamical system $(\xi, \mathfrak{F}_\xi, m, T_\xi)$[1] is exact. The discussion above tells us what could be a procedure for proving that a system is K-mixing:

a) find a partition $\xi$ of the space $X$ such that the induced map $T_\xi$ is exact;

b) prove that the partition $\xi$ of the first point is the 'special' partition that satisfy point a) and b) of Definition 6.1.1.

In the next section, we will apply such a procedure to our dynamical system which is a Lebesgue space. We anticipate that the partition $\xi^s$ into stable manifolds (yet to be defined) will take the place of $\xi$ in the preceding considerations. We will call $\mathfrak{B}_s$ the $\sigma$-algebra correspondent to $\xi^s$.

[1]To be precise, when one consider the natural measure on the elements of a measurable partition $\xi$, the correct notion is the one of *factor measure* $m_\xi$ which we will define later. Anyway, one can safely think of $m_\xi$ as $m$ 'evaluated' only on the elements of $\xi$.

## 6.2 Setup and result

Here we describe general conditions (H1)-(H6) under which our theorem holds. Most of them with minimal modifications were employed in [CM] to establish general conditions under which the *Hopf method* can be applied. Condition (H6) needs to be added to deal with infinite measure systems.

**(H1) Smoothness with singularities**. Suppose $\mathcal{M}$ is a Riemannian manifold (pheraps, with boundary and corners) of dimension $n$. Let $\mathcal{F}$ be a $C^r$ smooth ($r \geq 2$) invertible map defined everywhere except singularities such that the hypothesis of Oseledets theorem are satisfied. We assume that $\mathcal{F}$ preserves a $\sigma$-finite measure $\mu$ on $\mathcal{M}$, which is equivalent to the Lebesgue measure (i.e., has an a.e. positive density). Let us call $\mathfrak{A}$ the $\sigma$-algebra of measurable subsets of $\mathcal{M}$. We assume that the Lyapunov exponents exist on a set of full $\mu$-measure. Therefore, a stable and unstable directions are defined at almost every point $x \in \mathcal{M}$. We call these $E^s_x \subset \mathcal{T}_x\mathcal{M}$ and $E^u_x \subset \mathcal{T}_x\mathcal{M}$, respectively, where $\mathcal{T}_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ at $x$; we also set $d_s = \dim E^s_x$, $d_u = \dim E^u_x$, and $d = \dim \mathcal{T}_x\mathcal{M} = d_s + d_u$. We assume that both $d_s$ and $d_u$ are constant where they are defined.
Let $\mathcal{S}_1$ and $\mathcal{S}_{-1}$ be the singularity set for the map $\mathcal{F}$ and $\mathcal{F}^{-1}$. Let us define, inductively

$$\mathcal{S}_{n+1} = \mathcal{S}_n \cup \mathcal{F}^{-1}(\mathcal{S}_n), \quad \mathcal{S}_{-(n+1)} = \mathcal{S}_{-n} \cup \mathcal{F}(\mathcal{S}_{-n}).$$

Furthermore, let

$$\mathcal{S}_\infty = \cup_{n=0}^\infty \mathcal{S}_n, \quad \mathcal{S}_{-\infty} = \cup_{n=0}^\infty \mathcal{S}_{-n},$$

be the sets of points where *some future* and, respectively, *some past* iterate of $\mathcal{F}$ is singular.

We now define local stable and unstable manifold (LSUMs).

**Definition 6.2.1.** Given a point $x \in \mathcal{M} \setminus \mathcal{S}_{+(-)\infty}$, we define a local (un)stable manifold $W^{s(u)}$ for $\mathcal{F}$ at $x$ to be a $C^1$ topological disk containing $x$ in its interior, not intersecting $\mathcal{S}_{+(-)\infty}$, and such that:

a) $\mathcal{T}_x W^{s(u)}(x) = E^{s(u)}(x)$, where $\mathcal{T}_x W$ is the tangent space of the submanifold $W$ at $x$;

b) $\forall y \in W^{s(u)}(x)$, $d(\mathcal{F}^n x, \mathcal{F}^n y) \to 0$, as $n \to +(-)\infty$;

c) if $W^{s(u)}_0$ is another such manifold, then so is $W^{s(u)} \cap W^{s(u)}_0$.

Let $\| - \|$ be the Riemannian norm on $\mathcal{TM}$. The distance that appears in (2) is the one canonically associated with the norm $\| - \|$. Condition (3) is a sort of uniqueness property. One can see that $\mathcal{F}^{-1}W^u(x)$ is a LUM at $\mathcal{F}^{-1}x$ and $\mathcal{F}W^s(x)$ is a LSM at $\mathcal{F}x$. i.e.,

$$\mathcal{F}^{-1}W^u(x) \subseteq W^u(\mathcal{F}^{-1}x) \quad \mathcal{F}W^s(x) \subseteq W^s(\mathcal{F}x) \tag{6.2.1}$$

**(H2) Stable and unstable manifolds**. We assume that at $\mu$-almost every point $x \in \mathcal{H}$ there exist a local stable manifold $W^s(x)$ and a local unstable manifold $W^u(x)$. These manifolds can be represented as the graph of $C^1$ functions with bounded derivatives and hence are Lipschitz. We also ask two conditions on the diameter of any LUM $W^u(x)$ which are quite standard. Namely, for almost any $W^u$ we ask that

$$\lim_{n\to\infty} diam(\mathcal{F}^{-n}W^u(x)) = 0, \tag{6.2.2}$$

and

$$\lim_{n\to\infty} diam(\mathcal{F}^{-n}W^u(\mathcal{F}^n(x))) = 0, \tag{6.2.3}$$

e.g., we may ask that the size of LUMs is less or equal some chosen constant (see the method for building unstable manifolds in [L]). Finally, we assume that both homogeneous local stable and unstable manifolds form a partition of the phase space $\mathcal{M}$.

**Remark 6.2.2.** In (H2) we postulated the existence a.e. of (un)stable manifolds. Proving it is in general not a simple task, especially for infinite measure systems. Nevertheless, it was proved their existence under very mild assumptions on the map $\mathcal{F}$. We refer to [KS] for finite measure systems and [L1] for the infinite case. Besides assuming their existence we will postulate that they satisfy some convenient properties (see (H3), (H4) and (H5)), with the understanding that they have been proved in many interesting cases.

**(H3) Measurability**. Let $\mathcal{U} \subset \mathcal{M}$ be a sufficiently small open set (e.g., a ball). Then the partition of $\mathcal{U}$ into local unstable manifolds $W^u(x) \cap \mathcal{U}$, $x \in \mathcal{U}$, is measurable. The same is true for the partition of $\mathcal{U}$ into stable manifolds.[2]

Let $\mathcal{DF}$ the differential of the map $\mathcal{F}$. We denote by $\mathcal{D}_u\mathcal{F}$ the restriction of $\mathcal{DF}$ to the unstable subbundle made up of unstable subspaces at any point. Denote by $\mathcal{J}_u\mathcal{F}$ the Jacobian of $\mathcal{D}_u\mathcal{F}$. In the right coordinates, $\mathcal{J}_u\mathcal{F}$ can be represented as the determinant of the so called Jacobian matrix. Let $z^j$ $1 \le j \le k$ be a system of coordinates (a chart) for some unstable manifold $W^u$, i.e., $z^j = z^j(x)$ for $x \in W^u$. When considered as a function on $W^u$, $\mathcal{J}_u\mathcal{F}$ becomes a function of the $k$ coordinates $\bar{z} = (z^1, ..., z^k)$, i.e., $\mathcal{J}_u\mathcal{F} = \mathcal{J}_u\mathcal{F}(\bar{z})$. As a function of $\bar{z}$, $\mathcal{J}_u\mathcal{F}$ is at least $C^1$ since $\mathcal{F}$ is $C^r$ with $r \ge 2$ and the

[2] This property is clear once we define the stable and unstable manifold in the same way as in [CM] as '*what is left after the infinite cut form singularities*' but it is not obvioul from our definition of stable and unstable manifolds and must be added explicitly.

unstable manifolds are $C^1$ manifolds.

**(H4) Distortion**. We assume that there exists a global cross section $\widetilde{\mathcal{M}} \subset \mathcal{M}$ of positive measure such that the partition $\{\widetilde{\mathcal{M}}, \mathcal{M} \setminus \widetilde{\mathcal{M}}\}$ is coarser that the partition into unstable manifolds, i.e.,

$$\widetilde{\mathcal{M}} = \bigcup_{x \in \widetilde{\mathcal{M}}} W^u(x).$$

Let $\mathcal{F}_1$ be the return map to $\widetilde{\mathcal{M}}$. We can define the Jacobian $\mathcal{J}_u \mathcal{F}_1^{-1}(\bar{z})$ as a function of the coordinates $\bar{z}$ of any unstable manifold $W^u \subset \widetilde{\mathcal{M}}$ as in the previous discussion. We assume that there exists a constant $\widetilde{D}$ such that for $\mu$-almost every $W^u \subset \widetilde{\mathcal{M}}$ we have

$$\widetilde{D}^{-1} < \left| \frac{\mathcal{J}_u \mathcal{F}_1^{-n}(y)}{\mathcal{J}_u \mathcal{F}_1^{-n}(z)} \right| < \widetilde{D}, \quad \forall n, \tag{6.2.4}$$

for all $y, z \in W^u$.

**Remark 6.2.3.** A special case of assumption (H4) is when $\widetilde{\mathcal{M}} = \mathcal{M}$ and $\mathcal{F}_1 = \mathcal{F}$. In that case, (H4) means that for any unstable manifold the distortion is bounded.

Let $Leb_{W^u}$ be the sub-dimensional Lebesgue measure on the $k$-dimensional manifold $W^u$. Consider a subset $A$ of some unstable manifold $W^u(\mathcal{F}_1^n(x_0))$ of positive $Leb_{W^u(\mathcal{F}_1^n(x_0))}$ measure. The pre-image of $A$ under $\mathcal{F}_1^n$ will be a subset of $\mathcal{F}_1^{-n} W^u(\mathcal{F}_1^n(x_0)) \subseteq W^u(x_0)$. We then have

$$\int_{W^u(\mathcal{F}_1^n(x_0))} 1_{\mathcal{F}_1^n A}(x) |\mathcal{J}_u \mathcal{F}_1^{-n}(x)| dLeb_{W^u(\mathcal{F}_1^n(x_0))}(x) = \int_{\mathcal{F}_1^{-n} W^u(\mathcal{F}_1^n(x_0))} 1_A(x) dLeb_{W^u(x_0)}(x) \tag{6.2.5}$$

The term $|\mathcal{J}_u \mathcal{F}_1^{-n}|$ is the Jacobian of the map $\mathcal{F}_1$ and tells us how the volume of an infinitesimal cube in $W^u$ change under the application of $\mathcal{F}_1^{-n}$. In particular, from (6.2.5) e (6.2.4), for every $A \subseteq \mathcal{F}_1^{-n} W^u(\mathcal{F}_1^n(x_0))$ and $\mathcal{F}_1^n A \subseteq W^u(\mathcal{F}_1^n(x_0))$, we have that for every $n$

$$\begin{aligned} &\widetilde{D}^{-1} Leb_{W^u(x_0)}(A \cap \mathcal{F}_1^{-n} W^u(\mathcal{F}_1^n(x_0))) \leq Leb_{W^u(\mathcal{F}_1^n(x_0))}(\mathcal{F}_1^n A) \\ &\leq \widetilde{D} Leb_{W^u(x_0)}(A \cap \mathcal{F}_1^{-n} W^u(\mathcal{F}_1^n(x_0))) \end{aligned} \tag{6.2.6}$$

**(H5) Absolute continuity**. Let $V^1, V^2 \subset \mathcal{M}$ be two small nearby $d^u$-dimensional submanifolds uniformly transversal to stable subspaces $E^s_x$. Denote

$$V^i_* = \{x \in V^i : W^s(x) \cap V^{3-i} \neq \emptyset\}$$

for $i = 1, 2$ (the above intersection is either empty or it consists of a single point). We assume that the *holonomy* map defined by

$$h : V^1_* \to V^2_*, \quad x \mapsto W^s(x) \cap V^2$$

is absolutely continuous; i.e., it has a finite positive jacobian with respect to the internal Lebesgue measure on $V^1$ and $V^2$. Smilarly, the holonomy map defined by sliding along unstable manifolds is absolutely continuous.

A simple fact that is derived in [CM] from the assumption (H5) is the following Lemma. Let $V \subset \mathcal{U}$ be a $d_u$-dimensional submanifold uniformly transversal to stable subsapces $E^s_x$. Denote by $Leb_V$ the internal Lebesgue measure on $V$. Denote

$$V_* = \{x \in V : W^s(x) \text{ exists}\}.$$

**Lemma 6.2.4.** *For any measurable subset $B \subset V_*$ we have*

$$Leb_V(B) > 0 \Longleftrightarrow \mu(\bigcup_{x \in B} W^s(x)) > 0.$$

We will need also the following extra hypothesis to deal with infinite measure systems.

**(H6) Recurrence**. We assume that the dynamical system $(\mathcal{M}, \mathcal{F}, \mu)$ is conservative.

We state here our main result.

**Theorem 6.2.5.** *(Structure Theorem for Hyperbolic Maps) Let $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ be a map satisfying hypothesis (H1)-(H6). Then the map $\mathcal{F}$ admits an ergodic decomposition, i.e., there exists a finite or countable partition*

$$\mathcal{M} = \mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2 \cup \ldots \tag{6.2.7}$$

*such that*

*a) each $\mathcal{E}_i$ is $\mathcal{F}$-invariant, $\mu(\mathcal{E}_0) = 0$ and $\mu(\mathcal{E}_i) > 0$ for $i \geq 1$;*

*b) the restriction $\mathcal{F} : \mathcal{E}_i \to \mathcal{E}_i$ is ergodic for every $i \geq 1$.*
*Furthermore, for every $i \geq 1$ there is a finite partition*

$$\mathcal{E}_i = \mathcal{E}^0_i \cup \ldots \cup \mathcal{E}^{k_i - 1}_i \tag{6.2.8}$$

*such that*

*c) $\mathcal{F}(\mathcal{E}^j_i) = \mathcal{E}^{j+1}_i$ and $\mathcal{F}(\mathcal{E}^{k_i}_i) = \mathcal{E}^1_i$ (cyclic permutation);*

*d) the map $\mathcal{F}^{k_i} : \mathcal{E}^j_i \to \mathcal{E}^j_i$ is K-mixing.*

**Corollary 6.2.6.** *Let $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ be a map satisfying hypothesis (H1)-(H6). Assume that $\mathcal{F}^n$ is ergodic for any $n \geq 1$. Then $\mathcal{F}$ is K-mixing.*

*Proof.* By ergodicity there exists only one ergodic component, i.e., $\mathcal{M} = \mathcal{E}_1 \mod \mu$. Furthermore, by the Structure Theorem, the ergodicity of the powers of $\mathcal{F}$ imply that $\mathcal{E}_1 = \mathcal{E}^0_1 \mod \mu$. Hence $\mathcal{F}(\mathcal{E}^0_1) = \mathcal{E}^0_1$ is K-mixing. □

## 6.3 Proof of the Structure Theorem

In order to prove the previous result we will proceed by steps, to begin with we give the following definitions.

**Definition 6.3.1.** Let $(X, \mathfrak{F}, \mu)$ be a measure space. An atom $A \in \mathfrak{F}$ is a positive measure set such that for any measurable $B \subset A$ with $\mu(B) < \mu(A)$, $B$ is the empty set. We say that a set $D \in \mathfrak{F}$ is *infinitely divisible with respect to* $\mathfrak{F}$ if for any subset $\mathfrak{F} \ni E \subseteq D$ there exists $\mathfrak{F} \ni F \subset E$ with $\mu(F) < \mu(E)$.

**Definition 6.3.2.** Let $(X, \mathfrak{F}, \mu)$ be a measure space and let $\mathfrak{T}$ be a (mod 0 equivalence class) sub-$\sigma$-algebra. We say that $\mathfrak{T}$ is atomic if every positive measure set belonging to $\mathfrak{T}$ is composed mod 0 by atoms of $\mathfrak{T}$. Equivalently, $\mathfrak{T}$ is in correspondence with a partition $\xi$ composed of atoms only. If $\mathfrak{T}$ is not atomic we will say that it is *infinitely divisible*. In other words, $\mathfrak{F}$ is infinitely divisible if there exists one (and hence infinite) infinitely divisible sets.

Let $\mathfrak{T}_{B_s}$ be the tail $\sigma$-algebra of stable manifolds and $\mathfrak{I}$ be the sub-$\sigma$-algebra of mod 0 invariant sets

$$\mathfrak{T}_{B_s} = \bigcap_{n=0}^{\infty} \mathcal{F}^{-n}\mathfrak{B}_s, \tag{6.3.1}$$

$$\mathfrak{I} = \{A \in \mathfrak{F} : \mathcal{F}^{-1}A = A \text{ mod } 0\}. \tag{6.3.2}$$

Clearly $\mathfrak{I} \subset \mathfrak{T}_{\mathcal{B}_s}$. Indeed, for every $A \in \mathfrak{I}$ we have that $A$ is mod 0 a union of stable manifolds and $\mathcal{F}^{-n}A = A$ for every $n \in \mathbb{N}$. Another important fact is that $\mathfrak{T}_{\mathcal{B}_s}$ remains the same whether we build it with $\mathcal{F}$ or with $\mathcal{F}^n$ for any $n \in \mathbb{N}$.

**Lemma 6.3.3.** $\bigcap_{i=0}^{\infty} \mathcal{F}^{-i}\mathcal{B}_s = \bigcap_{i=0}^{\infty} (\mathcal{F}^n)^{-i}\mathcal{B}_s$

*Proof.* Indeed, the '$\subseteq$' direction is obvious and for any $N \in \mathbb{N}$. Furthermore,

$$\bigcap_{i=0}^{N} (\mathcal{F}^n)^{-i}\mathcal{B}_s = \mathcal{F}^{-n\times N}\mathcal{B}_s = \bigcap_{i=0}^{n\times N} \mathcal{F}^{-i}\mathcal{B}_s,$$

where we used the inclusion $\mathcal{B}_s \supseteq \mathcal{F}^{-1}(\mathcal{B}_s)$. Whence the other direction. $\square$

We will not prove the structure theorem directly, rather we will make use of the following result.

**Theorem 6.3.4.** *If $\mathfrak{T}_{B_s}$ is atomic then the dynamical system $(\mathcal{M}, \mathcal{F}, \mu)$ satisfies the statement of the Structure Theorem 6.2.5.*

*Proof.* Following the discussion of section (6.1) we first study the tail $\sigma$-algebra of the map $\mathcal{F}_{\xi^s}$ induced by the partition into stable manifolds $\xi^s$ and then prove that $\xi^s$ satisfy conditions a) and b) of Definition 6.1.1.
Since $\mathfrak{I} \subset \mathfrak{T}_{\mathcal{B}_s}$ and $\mathfrak{T}_{B_s}$ we have that the partition associated to $\mathfrak{T}_{\mathcal{B}_s}$ is finer than the one associated to $\mathfrak{I}$. Since the first of these two partitions is atomic, also the second is and so $\mathfrak{I}$ is an atomic $\sigma$-algebra and there are countable many ergodic components.
Now, let us consider the map $\mathcal{F}$ restricted to one of such ergodic components $\mathcal{E}_i$. By the atomicity of $\mathfrak{T}_{B_s}$ there must exists an atom $A \subset \mathcal{E}_i$ of positive measure. By the conservativity of $\mathcal{F}$ there exists a minimum $k \in \mathbb{N}^+$ such that $\mu(A \cap \mathcal{F}^k A) > 0$. This means that $\mathcal{F}^k A = A$ (mod 0) since $A$ is an atom of $\mathfrak{T}_{\mathcal{B}_s}$ and $\mathfrak{T}_{\mathcal{B}_s}$ is $\mathcal{F}$-invariant. Indeed, there are two possibilities; either $A \cap \mathcal{F}^k(A) \subsetneq A$ or $A \subsetneq \mathcal{F}^k(A)$ (both equations should be understood (mod 0)). But in the first case $A$ would not be an atom of $\mathfrak{T}_{\mathcal{B}_s}$ since $A \cap \mathcal{F}^k(A) \in \mathfrak{T}_{\mathcal{B}_s}$ and is (mod 0) strictly included in $A$. In the second case, one has $\mathcal{F}^{-k} A \subsetneq A$ and again $A$ would not be an atom of $\mathfrak{T}_{\mathcal{B}_s}$.
To match with the notation of theorem 6.2.5 we may call $A = \mathcal{E}_i^0$, $\mathcal{F}(A) = \mathcal{E}_i^1$, ... , $\mathcal{F}^{k-1}(A) = \mathcal{E}_i^{k-1}$. By the ergodicity of $\mathcal{F}$ restricted to $\mathcal{E}_i$ and the fact that $A$ has positive measure we have $\bigcup_{n=-\infty}^{+\infty} \mathcal{F}^n(A) = \bigcup_{n=0}^{k-1} \mathcal{F}^n(A) = \mathcal{E}_i$ (mod 0). Now, since every tail $\sigma$-algebra relative to $\mathcal{F}$ is equal to the tail $\sigma$-algebra relative to $\mathcal{F}^n$ $\forall n \in \mathbb{N}^+$ (see Lemma 6.3.3) we have that the tail $\sigma$-algebra of $\mathcal{F}^k$ restricted to $\mathcal{E}_i^j$ is precisely $\mathcal{E}_i^j$. This shows that the third property of k-mixing is satisfied by the map $\mathcal{F}^k_{|\mathcal{E}_i^j} : \mathcal{E}_i^j \to \mathcal{E}_i^j$.
Now, we prove the first two properties of K-mixing for the map $\mathcal{F}^k_{|\mathcal{E}_i^j}$ using the properties of the stable partition $\xi^s$. Denote by $\xi^s_{|\mathcal{E}_j^i} := \{W^s \cap \mathcal{E}_j^i | W^s \in \xi^s\}$. The fact that $\mathcal{F}\xi^s \geq \xi^s$ follows directly by the second of (6.2.1) and imply that $\mathcal{F}^k_{|\mathcal{E}_j^i} \xi^s_{|\mathcal{E}_j^i} \geq \xi^s_{|\mathcal{E}_j^i}$. The proof that $\bigvee_n \mathcal{F}^n \xi^s = \epsilon$ is included in the next lemma. From that and the fact that $\mathcal{F}^k_{|\mathcal{E}_j^i} \xi^s_{|\mathcal{E}_j^i} \geq \xi^s_{|\mathcal{E}_j^i}$ $\forall k > 0$, it follows that $\bigvee_n \left(\mathcal{F}^k_{|\mathcal{E}_i^j}\right)^n \xi^s_{|\mathcal{E}_i^j} = \epsilon_{|\mathcal{E}_i^j}$ (here $\epsilon_{|\mathcal{E}_i^j}$ is the partition of $\mathcal{E}_i^j$ into points), for all $i, j$ and therefore each $\mathcal{F}^k_{|\mathcal{E}_i^j}$ is K-mixing. □

**Lemma 6.3.5.** *If $\xi^s$ is the partition of $\mathcal{M}$ into stable manifolds for the map $\mathcal{F}$, then $\bigvee_n \mathcal{F}^n \xi^s = \epsilon$.*

*Proof.* Fix any $\epsilon \in \mathbb{R}^+$ and $N \in \mathbb{N}^+$ and let $W \in \xi_s$. Define

$$f_n(W) = diam(\mathcal{F}^n W);$$
$$A_{\epsilon,N} = \{W | \exists n \geq N : f_n(W) > \epsilon\}.$$

Let $B_\epsilon := \bigcap_{N=0}^{\infty} A_{\epsilon,N}$, note that $A_{\epsilon,N} \searrow B_\epsilon$. One has that $\mu(B_\epsilon) = 0$ for any $\epsilon > 0$. Indeed, if $W \in B_\epsilon$ then there would exist an infinite sequence of positive integer $\{n_1, n_2, ...\}$ such that $diam(\mathcal{F}^{n_k} W) > \epsilon$ for any $n_k$ belonging to that sequence, which is a contradiction with (6.2.2). Since for any stable manifold $W_1^s$ there exists another stable manifold $W_2^s$ such that $\mathcal{F}^{n+1} W_1^s \subseteq \mathcal{F}^n W_2^s$, we have that for any element of $\bigvee_n \mathcal{F}^n \xi^s$ that has diameter grater than $\epsilon$, there exists some $W$ belonging to $B_\epsilon$. But $\mu(B_\epsilon) = 0$, for all $\epsilon > 0$. Therefore, almost each element of $\bigvee_n \mathcal{F}^n \xi^s$ has zero diameter and hence is a point. $\square$

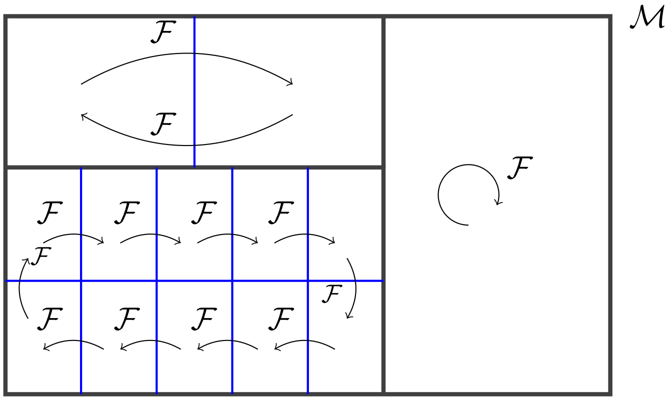


Figure 10: the atomic tail $\sigma$-algebra $\mathfrak{T}_{\mathcal{B}_s}$.

Let $\xi^u$ be the partition of $\mathcal{M}$ into unstable manifolds. Since $\xi^u$ is measurable (see (H3)) (both Lebesgue and $\mu$, since these measure are equivalent by (H1)), it is well defined a conditional measure on almost each unstable manifold, which we will denote by $\nu_{W^u}$. This measure has the important property of *disintegration*. Let $A$ be a measurable set with positive measure, then

$$Leb(A) = \int_{\xi^u} \nu_{W^u(x)}(W^u(x) \cap A) dLeb_{\xi^u}.$$

$Leb_\xi$ is called *factor measure* and we explain here how it is defined. If we call $\phi : \mathcal{M} \to \xi^u$ the natural projection which associates to each point in $\mathcal{M}$ the element of $\xi^u$ that contains that point, i.e., $\phi(x) = \xi(x)$, then $Leb_{\xi^u}$ takes value on the $\sigma$-algebra of subsets of $\xi^u$ defined by $\mathfrak{F}_{\xi^u} := \{D \subseteq \xi^u | \phi^{-1}(D) \in \mathfrak{F}\}$ and is defined by

$$Leb_{\xi^u}(D) = Leb(\phi^{-1}D). \tag{6.3.3}$$

Finally, we say that two measures $\lambda_1$ and $\lambda_2$ are equivalent if $\lambda_1$ is absolutely continuous with respect to $\lambda_2$ and vice-versa (in symbols $\lambda_1 \ll \lambda_2$ and $\lambda_2 \ll \lambda_1$) and we write $\lambda_1 \sim \lambda_2$. Such relation is an equivalence relation. Two equivalent measures have an almost everywhere positive density one with respect to the other.

**Lemma 6.3.6.** *Let $\xi^u$ be the partition into unstable manifolds. Once we disintegrate the Lebesgue measure over this partition, the conditional measure $\nu_{W^u}$ so obtained is equivalent to the internal Lebesgue measure $Leb_{W^u}$ supported on the unstable manifolds.*

*Proof.* The key property we use is the absolute continuity that we have assumed in (H5). Consider a $d_s$-dimensional hyperplane $h'$ and a small neighborhood $\mathcal{U}$. We can choose $\mathcal{U}$ small enough and $h'$ such that $h := h' \cap \mathcal{U}$ is transversal to the unstable direction for any $x \in h$. Set $A = \bigcup_{x \in h} W^u(x)$. Suppose $B \subset \mathcal{M}$ is measurable and $B \subset A$. Then, by disintegration

$$Leb(B) = \int_{\xi^u} \nu_{W^u(x)}(W^u(x) \cap B) dLeb_{\xi^u}(x). \tag{6.3.4}$$

By absolute continuity (see in particular Lemma 6.2.4) if $\mathfrak{g} \subset \xi^u \cap A$

$$Leb_{\xi^u}(\mathfrak{g}) = Leb(\bigcup_{W^u \in \mathfrak{g}} W^u) = 0 \Longleftrightarrow Leb_h(g) = 0, \tag{6.3.5}$$

where $g$ is the 'projection' of $\mathfrak{g}$ in $h$, i.e., $W^u(x) \in \mathfrak{g} \Leftrightarrow x \in g$ and $Leb_h$ is the standard $d_s$-dimensiona Lebesgue measure on the hyperplane $h$. We then have

$$Leb(B) = 0 \Longleftrightarrow \int_h \nu_{W^u(x)}(W^u(x) \cap B) dLeb_h(x) = 0. \tag{6.3.6}$$

On the other hand, if we let $r(x)$ be the $d_u$-dimensional hyperplane intersecting $h$ in $x$ and orthogonal to it, using Fubini theorem and absolute continuity one has that

$$Leb(B) = \int_h Leb_{r(x)}(r(x) \cap B) dLeb_h(x) = 0 \Longleftrightarrow \int_h Leb_{W^u(x)}(W^u(x) \cap B) dLeb_h(x) = 0. \tag{6.3.7}$$

By (6.3.6) and (6.3.7), for any measurable $B \subseteq A$

$$\int_h Leb_{W^u(x)}(W^u(x) \cap B) dLeb_h(x) = 0 \Leftrightarrow Leb(B) = 0 \Leftrightarrow \int_h \nu_{W^u(x)}(W^u(x) \cap B) dLeb_h(x) = 0. \tag{6.3.8}$$

Therefore,

$$\begin{aligned} &\text{For } Leb_h\text{-almost every } x \in h,\ Leb_{W^u(x)}(B \cap W^u(x)) = 0 \Longleftrightarrow \\ &\text{For } Leb_h\text{-almost every } x \in h,\ \nu_{W^u(x)}(B \cap W^u(x)) = 0 \end{aligned} \tag{6.3.9}$$

Hence, again by absolute continuity, for *Leb*-almost every $W^u \in A$ one has

$$Leb_{W^u}(W^u \cap B) \sim \nu_{W^u}(W^u \cap B), \tag{6.3.10}$$

for all measurable $B \subset A$. Now, varying the hyperplane $h$ and building the sets $A_h = \bigcup_{x \in h} W^u(x)$ we can cover all $\mathcal{M}$, so that (6.3.10) actually holds for any measurable $B \subset \mathcal{M}$.

$\square$

**Lemma 6.3.7.** *Let $A \subset \mathcal{M}$ be a positive measure set, i.e., $\mu(A) > 0$, and let $\xi^u \cap A$ be the partition of $A$ into unstable manifolds $W^u(z)$, $z \in A$. Then almost all point $x \in A$ is a density point with respect to the internal Lebesgue measure $Leb_{W^u(x)}$ of the submanifold $W^u(x)$.*

*Proof.* First we note that $\mu(A) > 0$ imply that $Leb(A) > 0$, since in (H1) we assumed that $\mu$ is equivalent to the Lebesgue measure on $\mathcal{M}$. Let $\widetilde{A}$ be the set of $Leb$-density points of $A$. Let us also define $B$ as

$$B = \{x \in A |\ x \text{ is not a density point of } W^u(x) \cap A \text{ with respect to } Leb_{W^u(x)}\}. \quad (6.3.11)$$

Similarly, let $\widetilde{B}$ be the set of $Leb$-density points of $B$. By the Lebesgue density point theorem $Leb(A \setminus \widetilde{A}) = Leb(B \setminus \widetilde{B}) = 0$. Then, to show that $Leb(B) = 0$ (which is the assertion of the theorem), we can just prove that $Leb(\widetilde{B}) = 0$. For this purpose, let $\bar{x}$ be a point belonging to $\widetilde{B}$, i.e.,

$$\lim_{\epsilon \to 0} \frac{Leb(B_\epsilon(\bar{x}) \cap B)}{Leb(B_\epsilon(\bar{x}))} = 1, \quad (6.3.12)$$

where $B_\epsilon(\bar{x})$ is the open ball of radius $\epsilon$ centered at $\bar{x}$. If we show that $\bar{x}$ is not a density point of $A$, the fact that $\widetilde{B} \subset A \setminus \widetilde{A}$ will then imply that $Leb(\widetilde{B}) = 0$. To understand whether $\bar{x}$ is a $Leb$-density point of $A$ we estimate the ratio $\frac{Leb(B_\epsilon(\bar{x}) \cap A)}{Leb(B_\epsilon(\bar{x}))}$ for smaller and smaller $\epsilon$. We notice that for all $\epsilon$

$$Leb_{W^u(x)}(B_\epsilon(\bar{x}) \cap A \cap B \cap W^u(x)) = 0, \quad (6.3.13)$$

since by definition the points of $B$ are not $Leb_{W^u(x)}$-density points for $W^u(x) \cap A$ and each positive measure set has a full measure subset of density points. Now, by Lemma 6.3.6, there exists an almost everywhere positive and bounded measurable function $L(x)$ such that for any measurable $C \subset \mathcal{M}$

$$\nu_{W^u}(W^u \cap C) = L(x) Leb_{W^u}(W^u \cap C).$$

Therefore

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{Leb(B_\epsilon(\bar{x}) \cap A)}{Leb(B_\epsilon(\bar{x}))} = \lim_{\epsilon \to 0} \frac{Leb(B_\epsilon(\bar{x}) \cap A \cap B)}{Leb(B_\epsilon(\bar{x}))} = \\ &= \lim_{\epsilon \to 0} \frac{1}{Leb(B_\epsilon(\bar{x}))} \int_{\xi^u} L(x) Leb_{W^u(x)}(B_\epsilon(\bar{x}) \cap A \cap B \cap W^u(x)) dLeb_{\xi^u} = 0. \end{aligned} \quad (6.3.14)$$

Hence $\bar{x}$ is not a $Leb$-density point for $A$.

□

The last auxiliary notion we need to establish is actually a consequence of our assumption about distortion.

**Lemma 6.3.8.** *Let $x_0 \in \widetilde{\mathcal{M}}$ be a density point of $W^u(x_0) \cap A$ with respect to $Leb_{W^u(x_0)}$. Then $\forall \epsilon > 0$, there exists a $\bar{k} \in \mathbb{N}$ sufficiently big such that $\forall k \geq \bar{k}$,*

$$Leb_{W^u(\mathcal{F}_1^k(x_0))}(\mathcal{F}_1^k A | W^u(\mathcal{F}_1^k(x_0))) > (1-\epsilon), \tag{6.3.15}$$

*where we have set $Leb(A|B) = Leb(A \cap B)/Leb(B)$.*

*Proof.* Since $\mathcal{F}_1^{-k}$ is continous for any $k$ on the unstable manifolds, $\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0))$ is a connected submanifold of $W^u(x_0)$ and in particular by (6.2.3)

$$\lim_{k\to\infty} diam(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0))) = 0.$$

Since $x_0$ is a density point for $A \cap W^u(x_0)$, one has

$$\lim_{k\to\infty} \frac{Leb_{W^u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0)) \cap A)}{Leb_{W^u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0)))} = 1.$$

In particular, for all $\epsilon > 0$ there exists a $\bar{k}$ sufficiently big such that for all $k \geq \bar{k}$, we have

$$Leb_{W^u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0)) \cap A) > (1-\epsilon) Leb_{W^u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k(x_0)))$$

Now, using equation (6.2.6), we have

$$\begin{aligned}
&Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0) \cap \mathcal{F}_1^k A) = Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0)) - Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0) \setminus \mathcal{F}_1^k A) \\
&= 1 - Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0) \setminus \mathcal{F}_1^k A) \geq 1 - \widetilde{D} Leb_{W^u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k x_0) \setminus A) \\
&> 1 - \epsilon \widetilde{D} Leb_{W_u(x_0)}(\mathcal{F}_1^{-k} W^u(\mathcal{F}_1^k x_0)) \geq 1 - \epsilon \widetilde{D}^2 Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0)) \\
&= Leb_{W^u(\mathcal{F}_1^k x_0)}(W^u(\mathcal{F}_1^k x_0))(1 - \widetilde{D}^2 \epsilon).
\end{aligned}$$

Note that $\mathcal{F}_1$ is invertible. Setting $\epsilon := \widetilde{D}^{-2}\epsilon$ at the beginning concludes the proof ($\widetilde{D}$ can be chosen independent of $\epsilon$). □

As we said, the sub-$\sigma$-algebra $\mathfrak{B}_s$ is the one in correspondence with the partition $\xi^s = \{W^s(x) : x \in \mathcal{M}\}$ into stable manifolds. Also the tail $\sigma$-algebra $\mathfrak{T}_{B_s}$ does correspond to a partition, which we denote by $\pi^s$, and is defined by the *meet* operation on partitions as ([CM])

$$\pi^s = \bigwedge_{m=0}^{\infty} \mathcal{F}^{-m} \xi^s.$$

The elements of $\pi^s$ are called *global stable manifolds* and we denote them with $W^s_G$. $W^s_G(x)$ is the element of $\pi^s$ containing $x$. We remind that $\mathcal{S}_\infty$ is the set of points for which $\mathcal{F}^n$ is discontinuous for some $n \in \mathbb{N}^+$. For any $x \in \mathcal{M} \setminus \mathcal{S}_\infty$,

$$W^s_G(x) = \bigcup_{n \geq 0} \mathcal{F}^{-n}(W^s(\mathcal{F}^n x)).$$

We note that even if the '*building blocks*' of $W^s_G(x)$, i.e., the stable manifolds $W^s(x)$, are contracting under the action of $\mathcal{F}$, $W^s_G(x)$ is invariant, because it gives information only about the asymptotic behavior of the system. Note, indeed, that the elements of $\mathfrak{T}_{B_s}$ must be (mod 0) a collection of these global stable manifolds. Thanks to Theorem 6.3.4 it is sufficient to show that the tail $\sigma$-algebra $\mathfrak{T}_{B_s}$ is atomic. The idea of the proof of the next theorem is to show that if we let $x$ vary in $\mathcal{M}$ and we consider all the positive measure sets belonging to $\mathfrak{T}_{B_s}$ which contain $x$, we have that their measure is everywhere bounded away from zero. Note that if we vary the point $x$ the measure of the tail-sets which contain $x$ can decrease to zero. Indeed, our goal is to prove that $\mathfrak{T}_{\mathcal{B}_s}$ is composed by atoms, but this does not exclude the possibility of having a sequence of atoms in $\mathfrak{T}_{\mathcal{B}_s}$ whose measures tend to zero. In order to prove atomicity, we will also employ the invariance property of global stable manifolds.

**Theorem 6.3.9.** $\mathfrak{T}_{B_s}$ *is atomic.*

*Proof.* Consider $A \in \mathfrak{T}_{B_s}$, $\mu(A) > 0$ and let $\widetilde{\mathcal{M}}$ be the global cross section introduced in assumption (H4). By (H2), through almost every $x$ pass an unstable manifold $W^u(x)$. By Lemma 6.3.7, almost every point $x$ in $A$ is a density point of $W^u(x) \cap A$ with respect to $Leb_{W^u(x)}$. Choose a $x_0 \in A$ with the aforementioned property. We observe that if $y \in W^u(x_0) \cap A$ then $W^s_G(y) \subset A$. Furthermore, for almost all $x \in \mathcal{M}$, there exists an $n(x) \in \mathbb{N}^+$ such that $\mathcal{F}^{-n(x)}x \in \widetilde{\mathcal{M}}$. Fix $c \in (0,1)$ and define

$$\begin{aligned} g(x) &:= \inf_{\substack{B \subset W^u(x) \\ Leb_{W^u(x)}(B) > c Leb_{W^u(x)}(W^u(x))}} \mu\Big(\bigcup_{y \in B} W^s(y)\Big); \\ \mathfrak{g}(x) &= g(\mathcal{F}^{-n(x)}x). \end{aligned} \tag{6.3.16}$$

Obviously $g$ is defined for $\mu$-a.e. $x \in \mathcal{M}$ and $g(x) > 0$ where it is defined. Indeed, if for some $B \subset W^u(x)$ it holds $\mu(\bigcup_{y \in B} W^s(y)) = 0$, we have (by Lemma 6.2.4) $Leb_{W^u(x)}(B) = 0 \leq c Leb_{W^u(x)}(W^u(x))$[3]. On the other hand, $g$ is not defined if for any $B \subset W^u(x)$, $Leb_{W^u(x)}(B) = 0$. But this can happen only if $W^u(x) = \{x\}$, a null-measure condition (in this case we simply say that in $x$ does not pass any unstable manifold). It follows

[3]To be precise, this holds only in the case in which for almost every $x \in W^u$ (with respect to the internal Lebesgue measure on $W^u$) does pass a stable manifold. But this happens for almost every unstable manifold $W^u$. One can see this using absolute continuity and the existence of stable and unstable manifold almost everywhere, see (H2) and (H5).

that $\mathfrak{g}$ is also defined almost everywhere and (where is defined) is positive.
Then, there exists $\delta \in \mathbb{R}^+$ such that $\mu(\{g(x) \geq \delta\}) > 0$. Consider the set

$$B := \{x \in \widetilde{\mathcal{M}} | g(x) \geq \min\{\mathfrak{g}(x_0), \delta\}\} = \{x \in \widetilde{\mathcal{M}} | g(x) \geq \min\{g(\mathcal{F}^{-n(x_0)}x_0), \delta\}\}, \tag{6.3.17}$$

we have that $\mu(B) > 0$. Obviously $\mathcal{F}^{-n(x_0)}x_0 \in B$ and, by the recurrence hypothesis, $\mathcal{F}^{-n(x_0)}x_0$ recurs infinitely often to $B$. Specifically, $\mathcal{F}^{n_k}(\mathcal{F}^{-n(x_0)}x_0) \in B \ \forall k \in \mathbb{N}$. Notice that $\mathcal{F}^{-n(x_0)}x_0$ is a density point of $W^u(\mathcal{F}^{-n(x_0)}x_0) \cap \mathcal{F}^{-n(x_0)}A$ with respect to $Leb_{W^u(\mathcal{F}^{-n(x_0)}x_0)}$, thanks to our choice of $x_0$. Call

$$\begin{aligned} A' &= \mathcal{F}^{-n(x_0)}A; \\ B \ni x_0' &= \mathcal{F}^{-n(x_0)}x_0. \end{aligned}$$

Consider any small $\epsilon > 0$, then by Lemma 6.3.8 there exists a sufficiently large $n_{\bar{k}}$ such that

$$\begin{aligned} &Leb_{W^u(\mathcal{F}^{n_{\bar{k}}}(x_0'))}(W^u(\mathcal{F}^{n_{\bar{k}}}(x_0')) \cap \mathcal{F}^{n_{\bar{k}}}(A') > \\ &> (1-\epsilon)Leb_{W^u(\mathcal{F}^{n_{\bar{k}}}(x_0'))}(W^u(\mathcal{F}^{n_{\bar{k}}}(x_0'))). \end{aligned} \tag{6.3.18}$$

From the fact that for all $n \in \mathbb{Z}$ $\mathcal{F}^n A \in \mathfrak{T}_{B_s}$, it follows that $\forall y \in W^u(\mathcal{F}^{n_{\bar{k}}}(x_0')) \cap \mathcal{F}^{n_{\bar{k}}}(A')$ one has $W^s(y) \subset W^s_G(y) \subset \mathcal{F}^{n_{\bar{k}}}(A')$. Hence

$$\bigcup_{y \in W^u(\mathcal{F}^{n_{\bar{k}}}(x_0)) \cap \mathcal{F}^{n_{\bar{k}}}(A')} W^s(y) \subset \mathcal{F}^{n_{\bar{k}}}(A'). \tag{6.3.19}$$

If we choose $1 - \epsilon > c$, by (6.3.18) and (6.3.16) one has

$$\begin{aligned} &\mu(A) = \mu(\mathcal{F}^{n_{\bar{k}}}(A')) \geq \mu(\bigcup_{y \in W^u(\mathcal{F}^{n_{\bar{k}}}(x_0')) \cap \mathcal{F}^{n_{\bar{k}}}(A')} W^s(y)) \geq \\ &g(\mathcal{F}^{n_{\bar{k}}}(x_0')) \geq \min\{g(x_0'), \delta\} = \min\{\mathfrak{g}(x_0), \delta\}, \end{aligned} \tag{6.3.20}$$

where in the first equality we used the invariance of $\mu$, in the second inequality we used (6.3.19), while the last two inequalities follow by the choice of $\epsilon$ and the fact that $\mathcal{F}^{n_{\bar{k}}}(x_0') \in B$ for any $k \in \mathbb{N}^+$, respectively. Hence, we have that for any positive measure set $A$ belonging to the tail $\sigma$-algebra $\mathfrak{T}_{B_s}$, it holds $\mu(A) \geq \min\{\mathfrak{g}(x_0), \delta\}$ for any $x_0$ density point of $W^u(x) \cap A$ with respect to $Leb_{W^u(x)}$. By Lemma 6.3.7 almost every point $x$ in $A$ has that property and so

$$\mu(A) \geq \min\{\delta, \operatorname*{ess\,sup}_{x \in A} \mathfrak{g}(x)\} = \min\{\delta, \sup_{a.e.\ x \in A} \mathfrak{g}(x)\}. \tag{6.3.21}$$

We are about to conclude the proof. For any $\sigma > 0$, in the level set $\{\mathfrak{g}(x) \geq \sigma\}$ $\mathfrak{T}_{B_s}$ has no infinitely divisible part.

Indeed, fix any some positive $\sigma$ and call $S = \{\mathcal{M} \setminus \bigcup_{A \text{ is an atom of } \mathfrak{T}_{\mathcal{B}_s}} A\}$. Note that S is infinitely divisible. If $\mu(S \cap \{\mathfrak{g}(x) \geq \sigma\}) > 0$, there exists an $\widetilde{x}$ belonging to such intersection such that any set in $\mathfrak{T}_{\mathcal{B}_s}$ which contains $\widetilde{x}$ has measure greater or equal $\sigma$ (this actually happens for almost any $x \in S \cap \{\mathfrak{g}(x) \geq \sigma\}$). Consider the set $\pi^s(\widetilde{x})$ of the partition $\pi^s$ corresponding to $\mathfrak{T}_{\mathcal{B}_s}$ containing $\widetilde{x}$. Now $\mu(\pi^s(\widetilde{x})) \geq \sigma$ and there cannot be any subset of $\pi^s(\widetilde{x})$ belonging to $\mathfrak{T}_{\mathcal{B}_s}$ of lower measure. Therefore, $S$ is not infinitely divisible with respect to $\mathfrak{T}_{\mathcal{B}_s}$ and it must be that $\mu(S \cap \{\mathfrak{g}(x) \geq \sigma\}) = 0$.
On the other hand,

$$\bigcup_{\sigma>0} \{\mathfrak{g}(x) \geq \sigma\} = \mathcal{M} \text{ mod } 0. \tag{6.3.22}$$

In fact, thanks to the hypothesis of $\sigma$-finiteness, there exists a nested sequence of sets $M_q$, $q \in \mathbb{N}^+$, increasing to $M$ (i.e., $M_1 \subset M_2 \subset ...$, $\bigcup_q M_q = \mathcal{M}$ ) such that $\mu(M_q) \in (0, \infty)$ $\forall q$. Fix any $q$ and consider $\{\mathfrak{g}|_{M_q}(x) \geq \sigma\} := C_{\sigma,q}$. We claim that

$$\mu(\bigcup_{\sigma>0} C_{\sigma,q}) = \mu(M_q). \tag{6.3.23}$$

If not, $\mu(\bigcup_{\sigma>0} C_{\sigma,q}) < \mu(M_q)$. Hence, on the set of positive measure $M_q \setminus \bigcup_{\sigma>0} C_{\sigma,q}$ we have $\mathfrak{g}(x) = 0$, which is absurd. This imply that

$$\mu(\bigcup_{\sigma>0} \Big\{\mathfrak{g}(x) \geq \sigma\Big\} \cap M_q) = \mu(M_q). \tag{6.3.24}$$

But this works for all $M_q$. □

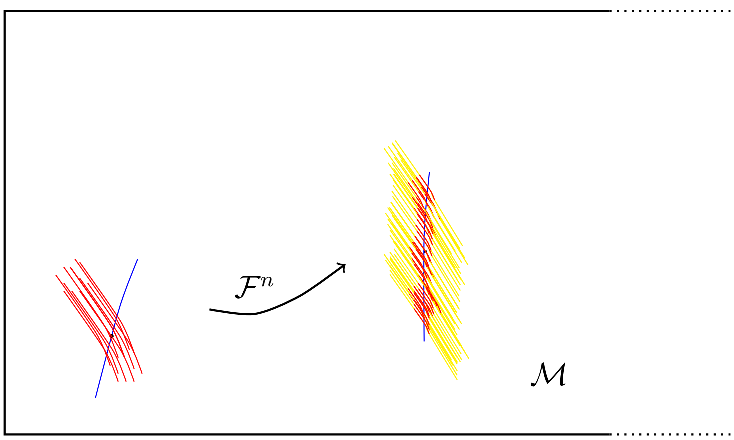


Figure 11: a schematic representation of the idea of the proof. The blue line is an unstable manifold $W^u$ and the black point is a $Leb_{W^u}$-density point for $A$, the red lines are stable manifolds that intersect $W^u$ in points belonging to $A$. The effect of $\mathcal{F}^n$ is to expand $W^u$ and contract the stable manifolds. Nevertheless, the global stable manifolds (some part of which being in yellow) are invariants.

## 6.4 Distortion

In this section we prove the hypothesis (H4) under more simple assumptions. We divide the assumptions that imply our *distortion bounds* (6.2.4) into four classes: assumptions on the LSUM manifolds (with the letter L), assumptions on the jacobian of the map $\mathcal{F}$ (with the letter J), assumptions on the singularity set (with the letter S) and hyperbolicity assumptions (with the letters Hy). In general, to prove hypothesis (H4) one needs one assumption from each of these four classes. We begin our analysis studying the case of $\mathcal{M}_1 = \mathcal{M}$ and $\mathcal{F}_1 = \mathcal{F}$ and we will generalize our results in the last section.

### 6.4.1 Assumptions for distortion

**Assumption on the LSUM**

For our applications, using *maximal* LSUM (i.e., the maximal sets of points for which properties of Definition 6.2.1 are verified) can be dangerous. Indeed, think at the two dimensional case of a planar Sinai billiard: the endpoints of some maximal, say, unstable manifold $W^u_{max}$ lay on $\mathcal{S}_{-\infty} = \bigcup_{i=0}^{\infty} \mathcal{F}^i \mathcal{S}_0$. This may have uncontrollable effects on the distortion of $W^u_{max}$. This is because, dynamical billiards have the property that the derivative of the Jacobian of the map (which is linked to distortion) blows up approaching $\mathcal{S}_0$ and points near $\mathcal{S}_{-\infty}$ happen to be near $\mathcal{S}_0$ infinitely often in the past. A possible solution to this problem is to assume the existence of LSUMs which are *homogeneous*, in the sense that are uniformly distant from the singularity set during all the time evolution. We now make a little digression about *invariant cone fields* to introduce the *unstable distance*, that we need to introduce *homogeneous* LSUMs.

**Definition 6.4.1.** Let $\mathcal{T}_x\mathcal{M} \simeq \mathbb{R}^d \simeq \mathbb{R}^{d_u} \oplus \mathbb{R}^{d_s}$ be the tangent space at $x \in \mathcal{M}$. And let

$$H_\rho = \{(u, v) \in \mathbb{R}^{d_u} \oplus \mathbb{R}^{d_s} | \|u\| \le \rho \|v\|\},$$
$$V_\rho = \{(u, v) \in \mathbb{R}^{d_u} \oplus \mathbb{R}^{d_s} | \|v\| \ \le \rho \|u\|\}$$

We call $H_\rho$ and $V_\rho$ the *horizontal and vertical $\rho$-cones* in $\mathcal{T}_x\mathcal{M}$, respectively. We define a *cone* $\mathcal{C}(x)$ in $\mathcal{T}_x\mathcal{M}$ the image of $H_\rho$ under any invertible linear transformation. A *cone field* is a measurable map that associates to each $x \in \mathcal{M}$ a cone $\mathcal{C}(x)$ in $\mathcal{T}_x\mathcal{M}$.

Given a cone field, one may be interested in the action of the differential $\mathcal{DF}$ on it. Of particular interest are the *invariant* cone fields.

**Definition 6.4.2.** A cone field $\mathcal{C}(x)$ is said to be invariant under the action of $\mathcal{F}$ if for almost every $x \in \mathcal{M}$

$$\mathcal{D}_x\mathcal{F}(\mathcal{C}(x)) \subseteq \mathcal{C}(\mathcal{F}x).$$

*Unstable cones* are invariant cones that approximate the unstable space $E^u_x$.

**Definition 6.4.3.** We call the cone filed $\mathcal{C}^u(x)$ *unstable cone field* if

$$E^u(x) = \bigcap_{n=0}^{\infty} \mathcal{D}_{\mathcal{F}^{-n}x}\mathcal{F}^n\mathcal{C}^u(\mathcal{F}^{-n}x).$$

We call a smooth curve $W$ unstable if its tangent space $\mathcal{T}_yW$ belongs to the unstable cone $\mathcal{C}^u(y)$ for every $y \in W$. Much like unstable cones are approximations of the unstable space at every point, the iterated images of any unstable curve $\mathcal{F}^n(W)$ for $n \in \mathbb{N}^+$ are approximations of (unions of) local unstable manifolds. Indeed, Pesin idea for the existence of an unstable manifold at some point was to consider smooth infinitesimal unstable curves in the past and evolve them obtaining further approximations of the unstable manifold at that point. Call $d^u$ the distance measured with the Riemannian norm $\| - \|$ along the shortest unstable curve connecting any two points (or a point with a set). Call $B^u(x, r)$ for $x \in \mathcal{M}$ and $r \in \mathbb{R}^+$ the ball centered at $x$ with radius $r$. Suppose that we want to build an homogeneous unstable manifold at some point $x \in \mathcal{M}$. Roughly speaking $W^u(x)$ is a connected set of points whose distance with $\mathcal{F}^{-n}x$ goes to zero as $n \to \infty$. We say that $W^u$ is also homogeneous if made up of points which stay at a 'safe' distance from the singularity set during their past dynamics.

**Definition 6.4.4.** We call the unstable manifold $W^u$ homogenous if there exists an $x' \in W^u$ such that

$$\mathcal{F}^{-n}y \in B^u(\mathcal{F}^{-n}x', d^u(\mathcal{F}^{-n}x', \mathcal{S}_{-1})/2),$$

for all $y \in W^u$ and for any $n \in \mathbb{N}^+$. Call the point $x'$ the *center* of the unstable manifold $W^u$. Note that it may happen that there are more than one point that satisfy the above requirement.

**(L1) Existence of homogenous LSUM**. We assume that the system of LSUM (whose existence is assumed in (H2)) is made by *homogeneous* LSUMs.

This leaves open the question about the conditions which guarantee the existence of a measurable partition of $\mathcal{M}$ into homogeneous LSUM. We will deal with this general question later.

### Assumptions on the jacobian

**(J) Bounds on the jacobian**. Given any fixed $W^u$, we assume that there exists a constant $D = D(W^u) \in \mathbb{R}^+$ possibly dependent on the unstable manifold and a $b > 0$ such that

$$\frac{\frac{\partial}{\partial z^j}\mathcal{J}_u\mathcal{F}(\bar{z})}{\mathcal{J}_u\mathcal{F}(\bar{z})} < \frac{D}{(d^u(x(\bar{z}), \mathcal{S}_{-1}))^b} \quad \forall 1 \leq j \leq k, \tag{6.4.1}$$

where $x(\bar{z})$ is the point on $W^u \subset \mathcal{M}$ correspondent to the coordinates $\bar{z}$.

**Assumptions on the singularity set**

**(S1) Neighborhood of the singularity lines**. We assume that for almost every $x \in \mathcal{M}$ the backward trajectory of $x$ does not approach the singularity set $\mathcal{S}_{-1}$ faster than a polynomial. In particulat, we assume that there exists a positive number $e$ and $Q(x) \in \mathbb{R}^+$ such that for almost every $x \in \mathcal{M}$

$$d^u(\mathcal{F}^{-i}x, \mathcal{S}_{-1}) \geq \frac{Q(x)}{(1+i)^e}, \tag{6.4.2}$$

for any $i \in \mathbb{N}^+$.

For $A \subseteq \mathcal{M}$ and $\epsilon > 0$ let us denote

$$A_{(\epsilon)} := \{x \in \mathcal{M} | d^u(x, A) \leq \epsilon\}. \tag{6.4.3}$$

**(S2) Neighborhood of singularity lines.** We assume that there exists a constant $C_1 \in \mathbb{R}^+$ and a positive $a$ such that for all $\epsilon > 0$ small enough

$$\mu(\mathcal{S}_{-1(\epsilon)}) < C_1 \epsilon^a. \tag{6.4.4}$$

We are interested in an application of our theorem where condition (S2) does not hold as we have stated before because the set $\mathcal{S}_{-1(\epsilon)}$ has infinite measure for every positive $\epsilon$ (this happens in Lorentz gas models). Therefore we introduce the slightly more general condition (S3) that implies (S2) and is satisfied by Lorentz gases, under which we will prove our general theorem. The 'philosophy' of (S3) is that even though the neighborhoods of the singularities of the entire dynamical system does occupy an infinite measure, we do not have to worry as long as (S2) holds for any finite measure subset of $\mathcal{M}$ and the portion of phase space accessible to a generic orbit (and so the opportunity to pass near some singularity line) grows not faster than polynomially in time.

**(S3) Neighborhood of singularity lines.** We assume that there exists a numerable partition $\{\mathcal{M}_p\}_{p \in \mathbb{N}}$ of $\mathcal{M}$ and $c > 0$ such that for every $p$ and $n \in \mathbb{N}$ there exists a subset $\mathcal{J} \subset \mathbb{N}$ with cardinality less or equal to $n^c$ such that

$$\mathcal{F}^{-n}\mathcal{M}_p \subseteq \bigcup_{i \in \mathcal{J}} \mathcal{M}_i, \quad \#\mathcal{J} \leq K(p) n^c, \quad \forall n, \tag{6.4.5}$$

where $K(p)$ is a positive constant. Furthermore, if we denote by $\mathcal{S}_{-1,p}$ the subset of $\mathcal{S}_{-1}$ belonging to $\mathcal{M}_p$, we assume that there exists a constant $C_1 \in \mathbb{R}^+$ and a positive $a$ such that for all $\epsilon > 0$ small enough

$$\mu(\mathcal{S}_{p,-1(\epsilon)}) < C_1 \epsilon^a. \tag{6.4.6}$$

**Remark 6.4.5.** When dealing with the Lorentz gas, we will identify the elements of the partition $\{\mathcal{M}_p\}_{p \in \mathbb{N}}$ with the phase space subsets correspondent to different scatterers.

**Assumptions on the hyperbolic behavior**

Suppose $\mathcal{F}$ is hyperbolic; i.e., there exists a full measure set $\mathcal{H} \subset \mathcal{M}$, $\mu(\mathcal{H} \setminus \mathcal{M}) = 0$, such that every point $x \in \mathcal{H}$ is hyperbolic (has no zero Lyapunov exponents). Then we have $T_x\mathcal{M} = E^u_x \oplus E^s_x$, where $E^u_x$ and $E^s_x$ denote the unstable and stable subspaces, respectively. We also assume that $d_u = dimE^u_x$ and $d_s = dimE^s_x$ are constant on $\mathcal{H}$, the spaces $E^u_x$ and $E^s_x$ depend continuously on $x \in \mathcal{H}$ and the differential of $\mathcal{F}$ leaves invariant this decomposition of the tangent bundle.

**(Hy1) Uniform hyperbolicity**. We remind that $\| - \|$ is the Riemannian norm on $\mathcal{TM}$. We assume that there exist $C \in \mathbb{R}^+$ and $\Lambda > 1$ such that for all $v \in E^u_x$,

$$\|D_x\mathcal{F}^{-n}v\|_{\mathcal{F}^{-n}x} \leq C\Lambda^{-n}\|v\|_x, \tag{6.4.7}$$

and for all $v \in E^s_x$,

$$\|D_x\mathcal{F}^{n}v\|_{\mathcal{F}^{n}x} \leq C\Lambda^{-n}\|v\|_x. \tag{6.4.8}$$

Alternatively, we may ask for the following, less demanding assumption.

**(Hy2) Non uniform hyperbolicity**. We assume that equations before holds but the constants $C$ and $\Lambda$ may depend on the unstable manifold $W^u$, i.e., $C = C(W^u)$ and $\Lambda = \Lambda(W^u)$. In this case, we still have that

$$diam(\mathcal{F}^{-n}W^u) \leq C(W^u)\Lambda(W^u)^{-n}diam(W^u). \tag{6.4.9}$$

### 6.4.2 Relations between assumptions

**Proposition 6.4.6.** *We have that* $(S2) \Rightarrow (S3) \Rightarrow (S1)$.

*Proof.* That $(S2) \Rightarrow (S3)$ is clear. Let us see that $(S3) \Rightarrow (S1)$.
Let $\mathcal{M}_q$ be an element of the partition $\{\mathcal{M}_p\}_{p\in\mathbb{N}}$ introduced in (S3) and consider the following sets

$$\begin{aligned} A_{i,q} &:= \left\{x \in \mathcal{M}_q \middle| d^u(\mathcal{F}^{-i}x, \mathcal{S}_{-1}) < \frac{1}{(i+1)^{2/(ac)}})\right\} = \mathcal{M}_q \cap \mathcal{F}^i\big((\mathcal{S}_{-1})_{[(i+1)^{-(2+c)/a}]}\big), \\ A_q &:= \left\{A_{i,q} \text{ i-infinitely often}\right\} = \bigcap_{m\in\mathbb{N}} \bigcup_{i\geq m} A_{i,q}. \end{aligned} \tag{6.4.10}$$

By the invariance of the measure and our fundamental assumption (H5”), we have

$$\mu(A_{i,q}) = \mu\big(\mathcal{F}^{-i}\mathcal{M}_q \cap (\mathcal{S}_{-1})_{[(i+1)^{-(2+c)/a}]}\big) \le \mu\big(\bigcup_{p\in\mathcal{J}} \mathcal{M}_p \cap (\mathcal{S}_{-1})_{[(i+1)^{-(2+c)/a}]}\big)$$
$$= \mu\big(\bigcup_{p\in\mathcal{J}} (\mathcal{S}_{p,-1} \cup \partial\mathcal{M}_p)_{[(i+1)^{-(2+c)/a}]}\big) \le \sum_{p\in\mathcal{J}} \mu\big((\mathcal{S}_{p,-1} \cup \partial\mathcal{M}_p)_{[(i+1)^{-(2+c)/a}]}\big) \le K(q)C_1 i^c (i+1)^{-2-c}, \tag{6.4.11}$$

where $\mathcal{J} \subset \mathbb{N}$ with $\#\mathcal{J} \le i^c$, by (S3). Therefore we can apply Borel-Cantelli lemma and conclude that $\mu(A_q) = 0$. This means that for almost every $x \in \mathcal{M}_q$ the backward trajectory $\{\mathcal{F}^{-i}x\}_{i\in\mathbb{N}}$ does not approach the singularity set faster than $1/(i+1)^{(2+c)/a}$. Since $\mathcal{M} = \bigcup_q \mathcal{M}_q$ and such union is countable, almost every $x \in \mathcal{M}$ has that property, too. In particular, for a.e. $x \in \mathcal{M}$ there exists a positive constant $Q = Q(x)$ such that

$$d^u(\mathcal{F}^{-i}x, \partial\mathcal{M} \cup \mathcal{S}_{-1}) \ge \frac{Q(x)}{(1+i)^{(2+c)/a}}.$$

Hence (6.4.2) holds with $e = (2+c)/a$. □

**Proposition 6.4.7.** $(Hy1) \Rightarrow (Hy2)$

*Proof.* Simply put $\Lambda(W^u) := \Lambda$ almost everywhere. □

### 6.4.3 Proving distortion estimates

**Theorem 6.4.8.** *Let $\mathcal{F}$ be a smooth hyperbolic map with singularities that preserves the measure $\mu$. Assume that $\mathcal{F}$ and its singularity set $\mathcal{S}_{-1}$ satisfy assumptions (L1), (J), (S1) (or (S2) by Proposition 6.4.6 ) and (Hy2). Then hypothesis (H4) on distortion bounds is satisfied by $\mathcal{F}$ (with $\widetilde{\mathcal{M}} = \mathcal{M}$).*

*Proof.* By (S1) for a.e. $x \in \mathcal{M}$ there exists a positive constant $Q = Q(x)$ and a positive number $e$ such that

$$d^u(\mathcal{F}^{-i}x, \mathcal{S}_{-1}) \ge \frac{Q(x)}{(1+i)^e}. \tag{6.4.12}$$

We claim that under assumptions (J) and (Hy2) there exists a constant $\widetilde{D}$ such that for $\mu$-almost every $W^u$ we have

$$\widetilde{D}^{-1} < \Big|\frac{\mathcal{J}_u\mathcal{F}^{-n}(y)}{\mathcal{J}_u\mathcal{F}^{-n}(z)}\Big| < \widetilde{D}, \quad \forall n,$$

for all $y, z \in W^u$. Indeed, let $y_i = \mathcal{F}^{-i}(y)$, $z_i = \mathcal{F}^{-i}(z)$, $\gamma_0$ a smooth curve in $W^u(\mathcal{F}^n(x_0))$ that links $y_0 = y$ with $z_0 = z$, $\gamma_i = \mathcal{F}^{-i}\gamma_0$ curves in $\mathcal{F}^{-i}W^u(\mathcal{F}^n(x_0))$ that link $y_i$ with

$z_i$, $\dot\gamma_i$ the unit tangent vectors to those curves. Taking the logarithm and applying the chain rule

$$\begin{aligned}
&\Big|\sum_{i=0}^{n}\log(\mathcal{J}_u\mathcal{F}^{-1}(y_i))-\log(\mathcal{J}_u\mathcal{F}^{-1}(z_i))\Big| \le \sum_{i=0}^{n}\Big|\log(\mathcal{J}_u\mathcal{F}^{-1}(y_i))-\log(\mathcal{J}_u\mathcal{F}^{-1}(z_i))\Big|\\
&\le \sum_{i=0}^{n}\big|\gamma_i\big|\sup_t\Big|\frac{\partial}{\partial\dot\gamma_i}\log(\mathcal{J}_u\mathcal{F}^{-1}(\gamma_i(t)))\Big| \le \sum_{i=0}^{n}|\gamma_0|C(W^u)\Lambda(W^u)^{-i}\sup_{\bar z\in Im(\gamma_i)} d_u \max_{1\le j\le d_u}\Big|\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}^{-1}(\bar z)\Big|\\
&\le \sum_{i=0}^{n}|\gamma_0|C(W^u)\Lambda(W^u)^{-i}d_u\sup_{\bar z\in Im(\gamma_i)}\Big|\frac{D}{(d^u(x(\bar z),\mathcal{S}_{-1}))^b}\Big|,
\end{aligned} \tag{6.4.13}$$

where $x(\bar z)$ is, as before, the point in $\mathcal{M}$ correspondent to the coordinate $\bar z$ on the unstable manifold we are dealing with. Furthermore, $|-|$ indicates both the absolute value of a real number and the length of a curve in the Riemannian norm $\|-\|$. In the last two inequalities, we have also used assumptions (Hy2) and (J). The computation above tells us that a sufficient condition for (6.2.4) to hold is the convergence of the series $\sum_{i=0}^{\infty}\Lambda(W^u)^{-i}\sup_{\bar z\in Im(\gamma_i)}\big|\frac{D}{(d^u(x(\bar z),\mathcal{S}_{-1}))^b}\big|$. We now show that such convergence takes place almost everywhere in $\mathcal{M}$. Indeed, since the inequality (6.4.12) holds almost everywhere, in almost every unstable manifold $W^u$ there exists at least one point[4] such that its backward orbit satisfy (6.4.12). Call such a representative point $x_{W^u}$,

$$d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1}) \ge \frac{Q(x_{W^u})}{(1+i)^e}. \tag{6.4.14}$$

By assumption (L1) one has that there exists a center $x'$ of $W^u$, i.e., a point $x'\in W^u$ such that

$$\sup_{\bar z\in Im(\gamma_i)} d^u(x(\bar z),\mathcal{F}^{-i}x') \le d^u(\mathcal{F}^{-i}x',\mathcal{S}_{-1})/2. \tag{6.4.15}$$

On the other hand, by (Hy2) the unstable distance between $x_{W^u}$ and the center $x'$ decreases exponentially with $\mathcal{F}^{-n}$, i.e.,

$$d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{F}^{-i}x') \le const(W^u)\times\Lambda(W^u)^{-i}. \tag{6.4.16}$$

[4]To be precise, *every* point of any unstable manifold does satisfy (6.4.12). This fact is indeed at the very foundation of the proof for the existence of unstable manifolds [CM]. Nevertheless, for our arguments it suffices to have just one point on almost all unstable manifolds with that property.

Using (6.4.15), (6.4.16) and triangular inequality repeatedly, we get

$$\inf_{\bar{z}\in Im(\gamma_i)} d^u(x(\bar{z}),\mathcal{S}_{-1}) \geq d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1}) - d^u(\mathcal{F}^{-i}x',\mathcal{F}^{-i}x_{W^u}) - \sup_{\bar{z}\in Im(\gamma_i)} d^u(x(\bar{z}),\mathcal{F}^{-i}x') \geq$$

$$\geq d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1}) - d^u(\mathcal{F}^{-i}x',\mathcal{F}^{-i}x_{W^u}) - \frac{1}{2}d^u(\mathcal{F}^{-i}x',\mathcal{S}_{-1}) \geq$$

$$\geq d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1}) - d^u(\mathcal{F}^{-i}x',\mathcal{F}^{-i}x_{W^u}) - \frac{1}{2}\Big(d^u(\mathcal{F}^{-i}x',\mathcal{F}^{-i}x_{W^u}) + d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1})\Big) =$$

$$= \frac{1}{2}d^u(\mathcal{F}^{-i}x_{W^u},\mathcal{S}_{-1}) - \frac{3}{2}d^u(\mathcal{F}^{-i}x',\mathcal{F}^{-i}x_{W^u}) \geq \frac{1}{2}\frac{Q(x_{W^u})}{(1+i)^e} - \frac{3}{2}\times const(W^u)\times \Lambda(W^u)^{-i}.$$

Therefore, for all $i$ big enough there exists a $Q_1 = Q_1(W^u)$ such that

$$\inf_{\bar{z}\in Im(\gamma_i)} d^u(x(\bar{z}),\mathcal{S}_{-1}) \geq \frac{Q_1(W^u)}{(1+i)^{e+1}}. \tag{6.4.17}$$

We note that $Q_1(W^u)$ depends only on the unstable manifold $W^u$ and not on the choice of $\gamma_0 \subset W^u$. Indeed, (L1) actually says that Equation (6.4.15) holds when we compute the infimum over all $W^u$ and not only over $\gamma_i$. Therefore, the last sum of (6.4.13) converges for almost every unstable manifold since the term $\sup_{\bar{z}\in Im(\gamma_i)}\left|\frac{D}{(d^u(x(\bar{z}),\mathcal{S}_{-1}))^b}\right|$ grows not faster than $i^{(e+1)b}$ for big $i$ almost everywhere. Hence, thanks to equation (6.4.13) we have that for every $n\in\mathbb{N}^+$, it holds

$$\sum_{i=0}^{n} |\gamma_0| C\Lambda^{-i} k \sup_{\bar{z}\in Im(\gamma_i)} \left|\frac{D}{(d^u(x(\bar{z}),\mathcal{S}_{-1}))^b}\right| \leq Q_2(W^u), \tag{6.4.18}$$

with $Q_2(W^u)$ being a positive constant which depends *only* on the unstable manifold $W^u(\mathcal{F}^n x_0)$ that one considers. Hence, if we set

$$\widetilde{D} = e^{Q_2(W^u)}$$

we have that the bound (6.2.4) holds on almost every $W^u$ with respect to $\mu$. Note that $\widetilde{D} = \widetilde{D}(W^u)$ but it does not depend on anything else. □

## 6.5 Distortion of the return map to a global cross section

Let $\mathcal{F}$ be a smooth maps with singularities that preserves the measure $\mu$. Let $\widetilde{\mathcal{M}} \subset \mathcal{M}$ be a global cross section for the map $\mathcal{F}$ such that $\partial\widetilde{\mathcal{M}}$ is piece-wise regular. Assume also that $\mathcal{F}$ has an hyperbolic structure, i.e., there exist almost everywhere homogeneous stable and unstable manifolds $\{W^s(x)\}_{x\in\mathcal{M}}$ and $\{W^u(x)\}_{x\in\mathcal{M}}$. We are now interested to establish distortion properties for the map $\mathcal{F}_1$. We now list our assumptions on the induced dynamical system $(\widetilde{\mathcal{M}}, \mathcal{F}_1, \mu)$. Let $\{A_n\}_{n\in\mathbb{N}^+}$ and $\{B_n\}_{n\in\mathbb{N}^+}$ two sequence of real numbers, we write $A_n \ll B_n$, $n \in \mathbb{N}^+$ if there exists a constant $D \in \mathbb{R}^+$ such that $A_n \leq DB_n$ for any $n$.

**(R) Assumption on the return times**. We assume that there exists a partition $\{\widetilde{\mathcal{M}}_q\}_{q\in\mathbb{N}^+}$ of $\widetilde{\mathcal{M}}$ with the following properties.

a)
$$\mathcal{F}_1^{-1}\widetilde{\mathcal{M}}_l \subseteq \bigcup_{q=1}^{l+1} \widetilde{\mathcal{M}}_q. \tag{6.5.1}$$

b) Let $x \in \widetilde{\mathcal{M}}$ and call $r = r(x)$ the past return time of $x$ to $\widetilde{\mathcal{M}}$. We assume that $r(x)$ is constant on each element of the partition $\{\widetilde{\mathcal{M}}_q\}_{q\in\mathbb{N}^+}$. Furthermore, if $r_q$ is the constant return time to $\widetilde{\mathcal{M}}$ defined in $\widetilde{\mathcal{M}}_q$, assume that
$$r_q \ll q^{\xi}, \tag{6.5.2}$$
for some positive $\xi$.

In other words, we ask that if we start with some $x \in \widetilde{\mathcal{M}}_p$ with past return time $r_p$, then $\mathcal{F}_1^{-1}(x)$ can be at most in $\widetilde{\mathcal{M}}_{p+1}$, its next past return time being $r_{p+1}$. Furthermore, with property b) we require $r_p$ grows at most like a polynomial as $p$ grows. Therefore, dynamics seen trough the eyes of the inverse return map interests only areas of the phase space which have past return times that grow polynomially.

Let $\widetilde{\mathcal{S}}_{-1} \subset \widetilde{\mathcal{M}}$ be the singularity set for the map $\mathcal{F}_1^{-1}$.

**Definition 6.5.1.** We call the partition $\{\widetilde{W}^u\}$ of $\mathcal{M}$ the partition of *pruned unstable manifolds* if for each $x \in \widetilde{\mathcal{M}}$ one has $\widetilde{W}^u(x) \subseteq W^u(x)$ and $\mathcal{F}_1^{-n}$ is smooth on $\widetilde{W}^u(x)$ for any $n$. In other words, we ask that for any $x \in \widetilde{\mathcal{M}}$ the pruned unstable manifold, $\widetilde{W}^u(x)$ does not intersect the set $\bigcup_{i=0}^{\infty} \mathcal{F}_1^{-i}\widetilde{\mathcal{S}}_{-1}$. For $x \in \mathcal{M} \setminus \widetilde{\mathcal{M}}$, we define the *pruned unstable manifold* at $x \in \mathcal{M} \setminus \widetilde{\mathcal{M}}$ as $\widetilde{W}^u(x) = \mathcal{F}^{-q}W^u(\mathcal{F}^q x)$ where $q$ is the smallest positive integer such that $\mathcal{F}^q x \in \widetilde{\mathcal{M}}$. Note that, since $\mathcal{F}^{-n}$ is smooth on $W^u(x)$ for any $n \in \mathbb{N}^+$, then we have also in this case that $\widetilde{W}^u(x) \subseteq W^u(x)$.

In analogy with Definition 6.4.4 we define *homogeneous* pruned unstable manifolds.

**Definition 6.5.2.** We call the unstable manifold $\widetilde{W}^u$ homogenous if there exists an $x' \in \widetilde{W}^u$ such that
$$\mathcal{F}^{-n}y \in B^u(\mathcal{F}^{-n}x', d^u(\mathcal{F}^{-n}x', \mathcal{S}_{-1} \cup \partial\widetilde{\mathcal{M}})/2),$$

for all $y \in \widetilde{W}^u$ and for any $n \in \mathbb{N}^+$. Call $x'$ *center* of the pruned unstable manifold $\widetilde{W}^u$. Note the presence of $\mathcal{F}^{-n}$ and not $\mathcal{F}_1^{-n}$ in the above definition.

In other words, we want that the past trajectories of points of $\widetilde{W}^u(x)$ stay sufficiently far both from the singularities of $\mathcal{F}^{-1}$ and to the 'additional' singularities of $\mathcal{F}_1^{-1}$ made by the the border of $\widetilde{\mathcal{M}}$

**(L2) Existence of pruned homogeneous LSUM**. We assume that there exists a system of pruned LSUM $\{\widetilde{W}^u\}$ which is made by *homogeneous* LSUM.

We deal later with the question about the existence of a partition of such LSUMs and the question about its measurability.

**(S4) Assumption on the singularity lines**. We assume that there exists a positive $\alpha_\sim$ such that

$$\mu((\partial\widetilde{\mathcal{M}})_{[\epsilon]}) \le const \times \epsilon^{\alpha_\sim}, \quad \forall \epsilon > 0. \tag{6.5.3}$$

**Lemma 6.5.3.** *Assume that the singularity set $\mathcal{S}_{-1}$ of $\mathcal{F}^{-1}$ satisfy hypothesis (S2). Let $r_p$ be the constant past return time defined on $\widetilde{\mathcal{M}}_p$, $p \in \mathbb{N}^+$. Then, there exists a positive $\alpha_1$ such that the set $(\widetilde{\mathcal{S}}_{-1} \cap \widetilde{\mathcal{M}}_p)$ for any $p$ has the property that*

$$\mu((\widetilde{\mathcal{S}}_{-1} \cap \widetilde{\mathcal{M}}_p)_{[\epsilon]}) \le r_p \times const \times \epsilon^{\alpha_1}, \quad \forall \epsilon > 0. \tag{6.5.4}$$

*Proof.* Fix any $p \in \mathbb{N}^+$. Then, $\widetilde{\mathcal{M}}_p \cap \widetilde{\mathcal{S}}_{-1} = \widetilde{\mathcal{M}}_p \cap (\mathcal{S}_{-r_p} \cup \partial\widetilde{\mathcal{M}})$. Now, if one sets $\alpha_1 = \min\{a, \alpha_\sim\}$, equation (6.5.4) follows from (S4), (S2) and the fact that $\mathcal{F}$ preserves the measure $\mu$. Indeed, we remind that $\mathcal{S}_{-r_p} = \bigcup_{i=0}^{r_p-1} \mathcal{F}^i\mathcal{S}_{-1}$. □

**(Hy3) Hyperbolicity assumption for the return map**. We assume that the return map $\mathcal{F}_1$ is (possibly non uniformly) hyperbolic, i.e., for any $\widetilde{W}^u \subset \widetilde{\mathcal{M}}$ there exist $C_1 = C_1(\widetilde{W}^u) \in \mathbb{R}^+$ and $\Lambda_1 = \Lambda_1(\widetilde{W}^u) > 1$ such that

$$diam(\mathcal{F}_1^{-n}(\widetilde{W}^u)) \le C_1\Lambda_1^{-n} diam(\widetilde{W}^u). \tag{6.5.5}$$

**(Hy4) Non-expansion assumption for the original map**. Furthermore, we assume that the map $\mathcal{F}^{-1}$ is non expanding on the pruned unstable manifolds $\widetilde{W}^u$, i.e., there exists a constant $\widetilde{C}(\widetilde{W}^u)$ such that for any $v \in \mathcal{T}_x\widetilde{W}^u$ we have

$$\|\mathcal{D}\mathcal{F}^{-n}(v)\|_{\mathcal{F}^{-n}x} \le \widetilde{C}(\widetilde{W}^u)\|v\|_x, \quad \forall n \in \mathbb{N}^+, \tag{6.5.6}$$

for any $x \in \widetilde{W}^u$.

**Proposition 6.5.4.** *The return map $\mathcal{F}_1 : \widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}$ preserves the measure $\mu$ (restricted to $\widetilde{\mathcal{M}}$).*

*Proof.* See Proposition 1.5.3, pag. 43 in [A]. □

**Proposition 6.5.5.** *Suppose that assumptions (R) and (S4) are satisfied. Furthermore, suppose that the singularity set $\mathcal{S}_{-1}$ for the map $\mathcal{F}^{-1}$ satisfy (S2). Then almost every $x \in \widetilde{\mathcal{M}}$ approach the singularity set $\widetilde{\mathcal{S}}_{-1}$ at most at polynomial speed in iterations of $\mathcal{F}_1^{-1}$. In other words, for almost all $x \in \widetilde{\mathcal{M}}$*

$$d^u(\mathcal{F}_1^{-i}x, \widetilde{\mathcal{S}}_{-1}) \geq \frac{const(x)}{(1+i)^\eta},$$

*for some positive $\eta$ and $i \in \mathbb{N}^+$.*

*Proof.* Fix any $p \in \mathbb{N}^+$ and consider all the trajectories going in the past starting at $\widetilde{\mathcal{M}}_p$. Now, consider the events

$$B_{n,p} := \mathcal{F}_1^n\big((\widetilde{\mathcal{S}}_{-1})_{[n^{-(2+(\xi+1))/\alpha_1}]}\big) \cap \widetilde{\mathcal{M}}_p = \mathcal{F}_1^n\big((\widetilde{\mathcal{S}}_{-1})_{[n^{-(2+(\xi+1))/\alpha_1}]} \cap \mathcal{F}_1^{-n}(\widetilde{\mathcal{M}}_p)\big).$$

The points in $B_{n,p}$ represents the past trajectories starting at $\widetilde{\mathcal{M}}_p$ which approach the singularity set $\widetilde{\mathcal{S}}_{-1}$ faster than $1/n^{-(2+(\xi+1))/\alpha_1}$ under iterations of $\mathcal{F}_1^{-n}$, $n \in \mathbb{N}^+$. By (6.5.1), one has that

$$\mathcal{F}_1^{-n}\widetilde{\mathcal{M}}_p \subseteq \bigcup_{l=1}^{p+n} \widetilde{\mathcal{M}}_l.$$

Hence, using (R) and Lemma 6.5.3

$$\mu(B_{n,p}) = \mu\big((\widetilde{\mathcal{S}}_{-1})_{[n^{-(2+(\xi+1))/\alpha_1}]} \cap \mathcal{F}_1^{-n}(\widetilde{\mathcal{M}}_p)\big) = \mu\big((\widetilde{S})_{[n^{-(2+(\xi+1))/\alpha_1}]} \cap \bigcup_{l=1}^{p+n} \widetilde{\mathcal{M}}_l\big) \leq$$

$$\leq \sum_{l=1}^{p+n} \mu\big((\widetilde{\mathcal{S}}_{-1})_{[n^{-(2+(\xi+1))/\alpha_1}]} \cap \widetilde{\mathcal{M}}_l\big) \leq \sum_{l=1}^{p+n} r_l \times const \times n^{-(2+(\xi+1))} \leq$$

$$\leq \sum_{l=1}^{p+n} l^\xi \times const \times n^{-(2+(\xi+1))} \leq (p+n) \times const \times (p+n)^\xi \times n^{-(2+(\xi+1))} \ll n^{-2}$$

Therefore the event $\bigcup_{m\geq 0}\bigcap_{n\geq m} B_{n,p}$ of $n$-infinitely many $B_{n,p}$ has zero measure by Borel-Cantelli Lemma. But the same holds for any element of the partition $\{\widetilde{\mathcal{M}}_p\}$, which is a countable partition. Setting $\eta = (2+(\xi+1))/\alpha_1$ concludes the proof. □

**Theorem 6.5.6.** *Let $\mathcal{F}$ be a map with an hyperbolic structure (i.e., existence a.e. of unstable manifolds $W^u$) that satisfy assumptions (J), (Hy4) and (S2). If assumptions (R), (L2), (S4) and (Hy3) are satisfied by $\mathcal{F}_1$ then hypothesis (H4) is verified for the dynamical system $(\widetilde{\mathcal{M}}, \mathcal{F}_1, \mu)$ endowed with the pruned unstable manifolds $\{\widetilde{W}^u\}$. In other words there exists a constant $\widetilde{D}$ such that for $\mu$-almost every $\widetilde{W}^u$ we have*

$$\widetilde{D}^{-1} < \Big|\frac{\mathcal{J}_u\mathcal{F}_1^{-n}(y)}{\mathcal{J}_u\mathcal{F}_1^{-n}(z)}\Big| < \widetilde{D}, \quad \forall n, \tag{6.5.7}$$

*for all $y, z \in \widetilde{W}^u$.*

*Proof.* The argument mirrors the proof of Theorem 6.4.8 with some modifications. Let $\widetilde{W}^u(\mathcal{F}^n x)$ be an hoogeneous pruned unstable manifold in $\widetilde{\mathcal{M}}$. Let $y_i = \mathcal{F}_1^{-i}(y)$, $z_i = \mathcal{F}_1^{-i}(z)$, $\gamma_0$ a smooth curve in $\widetilde{W}^u(\mathcal{F}_1^n(x_0))$ that links $y_0 = y$ with $z_0 = z$, $\gamma_i = \mathcal{F}_1^{-i}\gamma_0$ curves in $\mathcal{F}_1^{-i}\widetilde{W}^u(\mathcal{F}_1^n(x_0))$ that link $y_i$ with $z_i$, $\dot{\gamma}_i$ the unit tangent vectors to those curves. Now, to study the limit for $n \to \infty$ of the ratio (6.5.7) we use the chain rule with respect the return times to $\widetilde{\mathcal{M}}$. Indeed, for any $x \in \mathcal{M}$ for which the map $\mathcal{F}_1^{-l}$ are defined for any $l \in \mathbb{N}^+$, we have

$$\lim_{n\to\infty} \mathcal{J}_u\mathcal{F}_1^{-n}(x) = \lim_{n\to\infty} \mathcal{J}_u\mathcal{F}_1^{-1}(\mathcal{F}_1^{-n+1}x)...\mathcal{J}_u\mathcal{F}_1^{-1}(\mathcal{F}_1^{-1}x)\mathcal{J}_u\mathcal{F}_1^{-1}(x).$$

Doing so and applying the logarithm in (6.5.7), we get (using (Hy3))

$$\begin{aligned}
&\Big|\sum_{i=0}^{l-1} \log(\mathcal{J}_u\mathcal{F}_1^{-1}(y_i)) - \log(\mathcal{J}_u\mathcal{F}_1^{-1}(z_i))\Big| \le \sum_{i=0}^{l-1}\Big|\log(\mathcal{J}_u\mathcal{F}_1^{-1}(y_i)) - \log(\mathcal{J}_u\mathcal{F}_1^{-1}(z_i))\Big| \\
&\le \sum_{i=0}^{l-1} |\gamma_i| \sup_{\bar{z}\in Im(\gamma_i)} d_u \max_{1\le j\le d_u}\Big|\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}_1^{-1}(\bar{z})\Big| \\
&\le \sum_{i=0}^{l-1} C_1(\widetilde{W}^u)\Lambda_1(\widetilde{W}^u)^{-i} \sup_{\bar{z}\in Im(\gamma_i)} d_u \max_{1\le j\le d_u}\Big|\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}_1^{-1}(\bar{z})\Big|,
\end{aligned} \tag{6.5.8}$$

where we remind that $d_u := dim(E^u)$. Now, applying the chain rule to the term $\log(\mathcal{J}_u\mathcal{F}_1^{-1}(\bar{z}))$ we have

$$\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}_1^{-1}(\bar{z})) = \sum_{l=0}^{r_i-1}\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}^{-1}(\mathcal{F}^{-l}\bar{z})).$$

We see that this term can grow at most polynomially in $i$. Indeed, the number $r_i$ of terms in the last sums goes as a polynomial in $i$ by assumption (R). Furthermore, by the non-expansion assumption (Hy4) one has that

$$\sup_{\bar{z}\in Im(\gamma_i)}\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}^{-1}(\mathcal{F}^{-l}\bar{z})) \le d_u \times \widetilde{C}(\widetilde{W}) \sup_{\bar{z}\in Im(\gamma_l)}\frac{\partial}{\partial z^j}\log(\mathcal{J}_u\mathcal{F}^{-1}(\bar{z})), \tag{6.5.9}$$

where $\gamma_l := \mathcal{F}^{-l}(\gamma_i)$. Furthermore, by the combination of assumptions (J) and (S2) on the map $\mathcal{F}$ and the fact that $\widetilde{W}^u$ is homogeneous by assumption (L2) each of these terms is bounded by a polynomial in $l$ (the arguments are the same presented in the proof of Theorem 6.4.8). Therefore the a.e-convergence of the series is proved. □

## 6.6 Existence and measurability of a partition into homogeneous LSUMs

Now that we have given sufficient conditions for the distortion hypothesis (H4), we focus on the problem *existence* and *measurability* of a partition into stable and unstable manifolds, i.e., hypothesis (H2) and (H3) of the main theorem. Since for proving distortion we needed the existence of *homogenous* stable and unstable manifolds (see hypothesis (L1) and (L2) in the previous section), we would like to account both for measurability of the partition and homogeneity of the LSUMs.

### 6.6.1 Existence

The problem of the existence of LSUMs for infinite measure maps has been already considered in [L1]. There, it is proved the following theorem that we report.

**Theorem 6.6.1.** *(Lenci) Let $\mathcal{M}$ be a Riemannian manifold, embedded in $\mathbb{R}^N$, and $(\mathcal{M}, \mathcal{F}, \mu)$ an invertible, recurrent, dynamical system on it. Denote the discontinuity set of $\mathcal{F}$ by $\mathcal{S}$. Assume that for some $\zeta$, the following holds:*

*a) $\mu\big((\mathcal{S} \cup \partial\mathcal{M})_{[\epsilon]}\big) \ll \epsilon^\zeta$, for $\epsilon \to 0^+$.*

*b) There exists a continuous, invariant cone bundle $\mathcal{C}$, such that $\forall x \in \mathcal{M}$, $\bigcap_n \mathcal{C}_n(z) = E^u(z)$, a subspace of $\mathcal{T}_z\mathcal{M}$. $\mathcal{C}_n(z)$ is defined as in Definition 6.4.3.*

*c) There exists an increasing norm $\| - \|'$ for cone vectors, that is, $\forall z \in \mathcal{M} \setminus \mathcal{S}$, $\exists \kappa(z) > 1$ such that $\forall v \in \mathcal{C}(z)$, $\|\mathcal{D}_z\mathcal{F}v\|'_{\mathcal{F}z} \geq \kappa(z)\|v\|'_z$.*

*d) Let us denote by $H$ the set where the expansion factor $\kappa$ is not bounded away by $1$, i.e., $H := \{z | \exists z_n \to z, \kappa(z_n) \to 1\}$. Then $\mu(H_{[\epsilon]}) \ll \epsilon^\zeta$, when $\epsilon \to 0^+$.*

*e) Denote by $\| - \|$ the Riemannian norm on $\mathcal{TM}$, and taken two functions $0 < p \leq q$ such that $\forall z \in \mathcal{M} \setminus \mathcal{S}$,*

$$p(z)\| - \|'_z \leq \| - \|_z \leq q(z)\| - \|'_z,$$

*then $p$ is locally bounded below, and $q(z) \ll [d^u(z, \mathcal{S} \cup \partial\mathcal{M})]^{-\theta}$, where $\theta$ is some positive number.*

*Then, for $\mu - a.e.z$, the local unstable manifold $W^u(z)$ exists. Furthermore, let us take a $\mathcal{M}_0 \subseteq \mathcal{M}$, $\mu(\mathcal{M}_0) < \infty$, such that $(\mathcal{S} \cup \partial\mathcal{M})_{[\epsilon_0]} \subseteq \mathcal{M}_0$, for some $\epsilon_0 > 0$. Then the $W^u(z)$ are exponentially contracting with respect to the return times to $\mathcal{M}_0$. This means that, given a $z \in \mathcal{M}_0$ for which $W^u(z)$ exists, and denoted by $\{-n_k\}_{k\in\mathbb{N}}$ the sequence of its return times in the past, then $\exists C, \lambda > 0$ such that*

$$\forall w \in W^u, \quad |\mathcal{F}^{-n_k} w - \mathcal{F}^{-n_k} z| \leq C e^{-\lambda k}, \quad \text{as } k \to \infty.$$

**Remark 6.6.2.** A similar conclusion can be stated about the existence of stable manifolds at $\mu - a.e.$ point in $\mathcal{M}$, considering the time-reverse hypothesis, where necessary.

For the complete proof, we refer the reader to [L1] and here we limit to make some considerations. The fundamental idea in the proof of the previous result is to extend Pesin theory in the setting of an infinite measure recurrent map with singularities. To deal with singularities the idea is to consider unstable curves that lie in smaller and smaller neighborhoods of a given point in the past and then taking their pushforward under the map $\mathcal{F}$. If we choose the diameter of the neighborhood that contains the unstable curves in the past sufficiently small, we can be sure that these curves never hit the singularity set and also stay at a certain distance from it during all their time evolution.
By property (3) of Definition 6.2.1 it is clear that LSUM are not unique. On the other hand, for each $x$ there exists a *maximal*, say, LUM $W^u_{max}$ with the property that any LUM $W^u(x)$ that contains $x$ is a subset of $W^u_{max}$. We now report a method for constructing maximal LSUMs presented in [CM] for Sinai Billiards. Let $Q_{-n}(x)$ be the connected component of $\mathcal{M} \setminus \bigcup_{i=1}^n \mathcal{S}_{-i}$ that contains $x$. Define the maximal LUM at $x$ by

$$\begin{aligned} \widetilde{W^u_{max}(x)} &:= \bigcap_{k\geq 1} \overline{Q_{-k}(x)}, \\ W^u_{max}(x) &:= \widetilde{W^u_{max}(x)} \setminus \partial \widetilde{W^u_{max}(x)}, \end{aligned}$$

where $\overline{Q_{-k}(x)}$ is the closure of $Q_{-k}(x)$ and we removed the border to make $W^u_{max}(x)$ open submanifolds. Theorem 6.6.1 automatically guarantees that these submanifolds exists almost everywhere (i.e., $W^u_{max}(x) \neq \emptyset$ for $\mu$-almost every $x$); furthermore their definition imply that if we define as $\xi^u_{max}$ the partition such that $\xi^u_{max} = W^u_{max}(x)$ whenever $W^u_{max}(x) \neq \emptyset$ and $\xi^u_{max} = \{x\}$ otherwise, then $\xi^u_{max}$ is a measurable partition (an argument for measurability is given in the next section, when we will prove the measurability of the partition $\xi^u_{hom}$ which is an homogeneous refinement of $\xi^u_{max}$). Furthermore, in the setting of Theorem 6.6.1, the LSUMs are proved to be exponentially contracting with respect to the return times to a finite measure set $\mathcal{M}_0$, thus yielding the validity of assumption (Hy3) in the case of a finite measure global cross section.

**Existence of pruned unstable manifolds**

Let us now consider the issue of existence of pruned unstable manifold for the return map to a global cross-section, i.e., the validity of assumption (L2).

**Theorem 6.6.3.** *Assume that the dynamical system $(\mathcal{M}, \mathcal{F}, \mu)$ satisfy all hypothesis of Theorem 6.6.1. Furthermore, assume that $\widetilde{\mathcal{M}}$ is a global cross-section with piece-wise regular border $\partial\widetilde{\mathcal{M}}$ which satisfy assumption (S4) and that the return map $\mathcal{F}_1^{-1}$ satisfy assumption (R) on the return times. Then for $\mu-$a.e. $x \in \widetilde{\mathcal{M}}$ there exists an homogeneous pruned unstable manifold.*

*Proof.* We can actually apply Theorem 6.6.1 to the dynamical system $(\widetilde{\mathcal{M}}, \mathcal{F}_1, \mu)$, the only problem being that condition a) is not satisfied by the singularity set $\widetilde{\mathcal{S}}_{-1}$ of the return map $\mathcal{F}_1$. Analyzing the proof of Theorem 6.6.1 in [L1] one finds that the use of condition a) is to establish the fact that $\mu$-almost every point approach the singularity set at most at polynomial speed. Therefore, the statement of the theorem follows by Proposition 6.5.5. □

### 6.6.2 Measurability

Here we present a method for building homogeneous LUMs that automatically guarantees their measurability as a partition. For a detailed discussion of the next ideas for Sinai billiards we refer to chapter 5 of [CM]. The need for using homogeneous LSUMs instead of simple LSUMs comes form the fact that the Jacobian of the map (and its differential) blows up approaching the singularity set $\mathcal{S}_{-1}$ in the unstable distance $d^u$. Therefore, the idea is to substitute $\mathcal{S}_{-1}$ with an *extended singularity set* $\mathbb{S}_{-1}$ built with $\mathcal{S}_{-1}$ and an infinite number of other manifolds of the same dimension of $\mathcal{S}_{-1}$ that accumulate at $\mathcal{S}_{-1}$ in the $d^u$-distance (see the picture below). Let $k_0$ be big and positive natural number, we define

$$\mathbb{S}_{-1} = \bigcup_{k \geq k_0} \mathbb{S}_{-1}^k,$$

$$\mathbb{S}_{-1}^k = \{x \in \mathcal{M}; d^u(x, \mathcal{S}_{-1}) = const \times \frac{1}{k^\psi}\},$$

for some $\psi \in \mathbb{R}^+$ that we will fix later. We see that $\mathbb{S}_{-1}$ is a union of infinitely many submanifolds of the same dimension of $\mathcal{S}_{-1}$ (possibly with a lower degree of smoothness). Assume that for some $\zeta \in \mathbb{R}^+$

$$\mu((\mathcal{S}_{-1})_\epsilon) \ll \epsilon^\zeta, \quad \text{and} \quad \mu((\mathbb{S}_{-1}^k)_\epsilon) \ll \epsilon^\zeta. \tag{6.6.1}$$

The first one is a standard assumption on the singularity set of the map $\mathcal{F}^{-1}$ and the second is a consequence of the definition of $\mathbb{S}_{-1}^k$ in most of the cases. Let $\mathbb{H}_k$ be the set

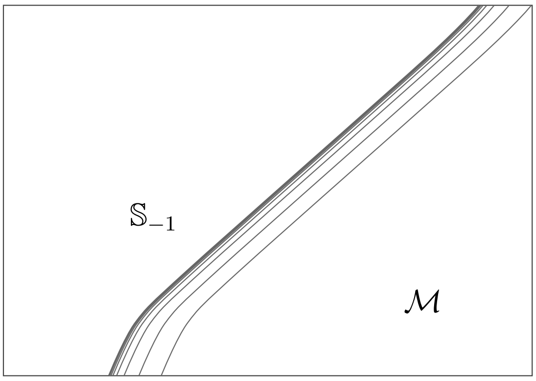


Figure 12: $\mathcal{S}_{-1}$ and other curves that accumulate at $\mathcal{S}_{-1}$ form the extended singularity set $\mathbb{S}_{-1}$.

of point which lies in between the two submanifolds $\mathbb{S}_k$ and $\mathbb{S}_{k+1}$, i.e.,

$$\mathbb{H}_k = \{x \in \mathcal{M} : \frac{1}{k^\psi} < d^u(x, \mathcal{S}_{-1}) < \frac{1}{(k+1)^\psi}\}.$$

We call $\mathbb{H}_k$ the *homogeneity strip of order k*.

Now consider two points $x$ and $y$ that lie in the same connected component of $\mathcal{M} \setminus (\mathbb{S}_{-1} \cup \mathcal{F}\mathbb{S}_{-1})$. We know that the unstable distance of $x$ and $y$ to the singularity set $\mathcal{S}_{-1}$ is controlled by the fact that $x$ and $y$ lie in the same connected component of $\mathcal{M} \setminus \mathbb{S}_{-1}$; moreover the unstable distance of $\mathcal{F}^{-1}x$ and $\mathcal{F}^{-1}y$ to the singularity set $\mathcal{S}_{-1}$ is controlled by the fact that $x$ and $y$ lie in the same connected component of $\mathcal{M} \setminus \mathcal{F}(\mathbb{S}_{-1})$. Therefore, points that lie in the same connected component of $\mathbb{Q}_{-k} \in \mathcal{M} \setminus (\mathbb{S}_{-1} \cup \mathcal{F}(\mathbb{S}_{-1}) \cup \mathcal{F}^2(\mathbb{S}_{-1}) \cup ... \cup \mathcal{F}^k(\mathbb{S}_{-1}))$ are *homogeneous* in the sense that their $d^u$-distance to the singularity set $\mathcal{S}_{-1}$ is bounded from below by a certain constant dependent on $\mathbb{Q}_{-k}$ *along all the past dynamics until time -k.* The idea now is to define homogeneous LUMs at some $x$ by the intersection of the closures of the connected components $\mathbb{Q}_{-k}(x)$ that contain $x$, i.e.,

$$\widetilde{W^u_{hom}(x)} := \bigcap_{k \geq 1} \overline{\mathbb{Q}_{-k}(x)},$$
$$W^u_{hom}(x) := \widetilde{W^u_{hom}(x)} \setminus \partial \widetilde{W^u_{hom}(x)}.$$

We remove the boundary points in the definition of $W^u_{hom}(x)$ because we would like it to be an open submanifold. This automatically guarantees a bound on the speed trough at which points in $W^u_{hom}(x)$ approach the singularity set $\mathcal{S}_{-1}$ under the action of $\mathcal{F}^{-1}$. The advantage of this method is that it guarantees immediately that the set $\{W^u_{hom}(x)\}_{x \in \mathcal{M}}$ is a partition and that is a measurable one. Precisely, define $\xi^u_{hom}$ as the partition of $\mathcal{M}$ such that $\xi^u_{hom}(x) = W^u_{hom}(x)$ if the latter is not empty, and $\xi^u(x) = \{x\}$ otherwise. To prove that $\xi^u_{hom}$ is measurable we need to construct a countable generator $\{B_m\}_{m=1}^{\infty}$ for it. Let $\{B_m\}_{m=1}^{\infty}$ denote the (countable) collection of $\{\overline{\mathbb{Q}_{-n}(x)}\}_{n=1,x\in\mathcal{M}}^{\infty}$ and let us prove that $\{B_m\}_{m=1}^{\infty}$ is indeed a generator for $\xi^u$. First, for every element $C \in \xi^u_{hom}$ and every $B_m$ either $C \subset B_m$ or $C \cap B_m = \emptyset$. This follows from the fact that for every fixed $x$ $\overline{\mathbb{Q}_{-n-1}(x)} \subseteq \overline{\mathbb{Q}_{-n}(x)}$ and for every fixed $n$ the sets $\{\mathbb{Q}_{-n}(x)\}_{x\in\mathcal{M}}$ form a partition of $\mathcal{M}$. Second, for every two distinct elements $C_1 \neq C_2$ of the partition $\xi^u$ there exists a $B_m$ such that $C_1 \subset B_m$ and $C_2 \cap B_m = \emptyset$. Indeed, if it were not so, both $C_1$ and $C_2$ would have to belong to the intersection $\bigcap_{k\geq 1}\overline{\mathbb{Q}_{-k}(x)}$ for the same sequence of components $\overline{\mathbb{Q}_{-k}(x)}$, but this would imply that are the same element, i.e., $C_1 = C_2$. This simply proves that the partition $\xi^u$ is measurable. On the other hand, this approach leaves open the question about the existence of $W^u_{hom}(x)$ at almost every point (i.e., the intersection $\bigcap_{k\geq 1}\overline{\mathbb{Q}_{-k}(x)}$ stabilizes at $\mu$-almost everywhere or we might end up with points?) and about the properties of $W^u_{hom}(x)$ (is it true that $W^u_{hom}(x)$ satisfies all the nice properties of Definition 6.2.1?). As concerns the existence of *bona fide* LUMs, we state the following Theorem.

**Theorem 6.6.4.** *Let $(\mathcal{M}, \mu, \mathcal{F})$ satisfy all hypothesis of Theorem 6.6.1. Then there exists some $\psi \in \mathbb{R}^+$ in the definition of $\mathbb{S}_{-1}$ such that for $\mu$-almost every point there exists $W^u_{hom}(x)$, i.e., $diam(W^u_{hom}(x)) > 0$ for $\mu$-almost every $x \in \mathcal{M}$.*

*Proof.* From the fact that for each $k \in \mathbb{N}^+$ the partition $\mathbb{Q}_k(x)$ is a refinement of the partition $Q_k(x)$, it follows that the partition $\xi^u_{hom}$ is a refinement of the partition $\xi^u_{max}$. Therefore, almost each $W^u_{max}$ is partitioned in other homogenous LUMs $W^{u,\alpha}_{max}$, i.e., $W^u_{max}(x) = \bigcup_{\alpha\in\mathcal{I}} W^{u,\alpha}_{hom}$, where $\mathcal{I} = \mathcal{I}(x)$ is a possibly continuous index set (at the end of the proof, we will see that $\mathcal{I}$ must be a subset of $\mathbb{N}$ for almost every $x$). Suppose that the converse of the statement of the theorem holds. Then, there would exists a measurable set $A \subset \mathcal{M}$ with the property that for any $x \in A$ one has that $\xi^u_{hom} = \{x\}$. Therefore, by Fubini theorem and absolute continuity of the partition $\xi^u_{max}$, there would exists a positive measure of maximal LUMs that intersect $A$ in a set of positive $Leb_{W^u}$-measure. We now show that there are choices of $\psi$ in Equation (6.6.2) for which this is impossible. Consider any typical $W^u_{max}$ with the aforementioned property of intersecting $A$ in a positive $Leb_{W^u}$-measure set. This may happen only if $\mathcal{F}^{-i}(W^u_{max})$ intersect the set $\mathbb{S}_{-1}$ infinitely often in $i$. Note that $\mathcal{F}^{-i}W^u_{max}$ is contracted by a factor $\Lambda^i$, $\Lambda > 1$ with respect to $W^u_{max}$ (if the map is not uniformly hyperbolic this can be done considering $\mathcal{F}$ as the return map to some global cross section, at least in all the examples considered above).

On the other hand, by the first one of hypothesis (6.6.1) and using again a standard Borel-Cantelli argument, one has that almost every $x \in \mathcal{M}$ approach $\mathcal{S}_{-1}$ at most at polynomial speed with exponent $2/\zeta$, i.e., for almost every $x$ there exists a constant possibly dependent on $x$ such that

$$d^u(\mathcal{F}^{-i}x, \mathcal{S}_{-1}) > \frac{const(x)}{(1+i)^{2/\zeta}}. \tag{6.6.2}$$

Therefore, choosing e.g. $\psi = 2/\zeta + 1$, for almost every $x$ the number of strips of $\mathbb{S}_{-1}$ accessible to its past dynamics increase as a polynomial in time. On the other hand, for $\mathcal{F}^{-i}(W_{max})$ being cut, one has that $\mathcal{F}^{-i}x_W$ must lie in a *const* $\times \Lambda^{-i}$ neighborhood of some $\mathbb{S}^k_{-1}$, whose measure decrease like an exponential by the second of (6.6.1). Hence, by a Borel-Cantelli argument this can happen only finitely many times for almost any $W^u_{max}$[5].

□

Finally, as far as the properties enjoyed by $W^u_{hom}(x)$ are concerned, in the Sinai-billiard case, it can be proved that they are smooth curves which enjoy all properties of Definition 6.2.1 and such that $\lim_{n\to\infty} \mathcal{F}^{-n} diam(W^u_{hom}(x)) = 0$ and $\lim_{n\to\infty} \mathcal{F}^{-n} diam(W^u_{hom}(\mathcal{F}^n x)) = 0$ a.e., as required in (H3) (again see [CM] chapter 4 and 5). We believe that these 'characteristic' properties can be proved also in our setting (i.e., under the hypothesis of Theorem 6.6.1).

## 6.7 Applications

In this final section we outline some possible applications of the main theorem 6.2.5.

### 6.7.1 Lorentz Gases and Tubes

For Periodic Lorentz Gases (PLG) with an hyperbolic structure recurrence and ergodicity are well known facts[6] (see [Sch] and [Co]), that readily imply the K-mixing for the return map $T_\alpha$ to the phase space subset $\mathcal{M}_\alpha$ associated to every scatterer (see [CM]) and the ergodicity of $T^n$ for $n \geq 1$. Corollary 6.2.6 imply that the map $T$ itself is K-mixing. The same holds for the class $\chi$ of aperiodic LG introduced in [L3], which is proved to be reccurrent. In that work it is proved that billiards belonging to $\chi$ enjoy an hyperbolic structure and are ergodic (in addition to be recurrent); thanks to Corollary 6.2.6 this is sufficient to prove that they are K-mixing as well. Examples of aperiodic LG belonging to the class $\chi$ are the *finite modifications* of PLG, introduced in [L3] as well. Our results apply also in the setting of quenched dynamical systems [CLS], even in dimensions greater than two [SLEC]. In those two works, using the backbone results of [L3] that guarantee the existence of a hyperbolic structure for almost every realization of the quenched dynamical system, one is able to prove recurrence for quenched Lorentz Tubes. Again, thanks to Corollary 6.2.6, this imply that the dynamics is K-mixing as well. Note that for Lorentz Gases we have mentioned so far hypothesis (J), (S3) and

[5]To be precise this Borel-Cantelli argument is not applicable since the constant in (6.6.2) depends on $x$. At any rate, this is not a real obstruction to our argument. Indeed one can partition $\mathcal{M}$ into a countable union of disjoint sets where such a constant is bounded from below and apply to each such set the aforementioned Borel-Cantelli argument.

[6]For a Lorentz gas at least, recurrence imply ergodicity (see [L3]).

(Hy1) are verified (see [L3], once again), thus leading to the validity of the distortion hypothesis (H4) trough Theorem 6.4.8.

### 6.7.2 Infinite cusp billiards

The last application we consider concerns the class of *infinite cusp billiards* studied by Lenci. In [L1] it is proved recurrence and the existence of an hyperbolic structure for this interesting class of infinite measure and non-uniformly hyperbolic dynamical system, as well as absolute continuity.[7] Furthermore, assumptions (R), (L2), (S4), (Hy3) and (Hy4) are verified for a map $T_1$ individuated by a suitable cross-section obtained placing transparent walls in the billiard table (for the relevant definitions and theorems we refer to [L1]), proving so the distortion hypothesis (H4), thanks to Theorem 6.5.6. Therefore, Corollary 6.2.6 actually implies that the billiard map $T$ is K-mixing. This last application motivates the effort made to formulate a theorem which requires the existence of distortion bounds only for the return map to a suitable cross-section (in the class of cusp billiards mentioned above, distortion bounds may not hold for the simple billiard map $T$).

[7]For Lorentz Gases the absolute continuity can be derived by the works of Sinai, see [CM].

# Ringraziamenti

Desidero innanzitutto ringraziare il Prof. Marco Lenci, per tutto quello che mi ha insegnato, per i molti dialoghi che abbiamo avuto e per essere sempre riuscito a trovare del tempo per me. Mi ritengo estremamente fortunato per tutto ciò. Desidero poi ringraziare la mia famiglia: i miei genitori e mio fratello. Non ho mai sentito mancare il vostro sostegno. Ringrazio tutti i miei amici fisici, specialmente tutti *i Ragazzi di via Mascarella* per questi 5 anni *intensi* che abbiamo passato assieme. Ringrazio anche i fisici 'grandi' (in particolare Ema, Simone, Sunny, Octavio, Maro, Cate, Michelangelo, Paolo, Marina, Federico e spero di non dimenticare nessuno...), perché mi sono sempre sentito accolto tra voi e siete ora tutti miei grandi amici. Passando ai matematici, vorrei ringraziare Umberto e Diego per avermi dato una mano in più occasioni e per le belle discussioni che abbiamo fatto. Un grazie particolare a Tabby per tutto quello di cui abbiamo parlato e per i suoi consigli, con lui vorrei anche scusarmi per aver insistentemente provato a farlo passare ai biliardi. Ringrazio anche i miei amici dello storico gruppo *Che cosa mangio*, per non avermi dimenticato nei lunghi periodi di assenza dovuta allo studio. Ringrazio Isabella e Alessia, perché non ci siamo mai persi di vista. Un ultimo ringraziamento va a Marco, Andrea e Isa, che in questi anni mi sono sempre stati vicino.

# Bibliography

[A] *Aaronson, J.: An introduction to infinite ergodic theory. Mathematical Surveys and Monographs, 50. Providence, RI: Amer. Math. Soc., 1997.*

[B] *A. Brown, Entropy, Lyapunov exponents, and rigidity of group actions. arXiv:1809.09192v4 [math.DS], 11 Jun 2019.*

[BL] *Bussolari, Luca and Lenci, Marco. (2008). Hyperbolic billiards with nearly flat focusing boundaries, I. Physica D: Nonlinear Phenomena. 237. 2272-2281. 10.1016/j.physd.2008.02.006.*

[BW] *Bungert, Leon and Wacker, Philipp. (2020). The lion in the attic – A resolution of the Borel–Kolmogorov paradox.*

[CK] *V. Climenhaga and A. Katok. Measure theory through dynamical eyes. Available at arxiv.org/abs/1208.4550.*

[CLS] *G. Cristadoro, M. Lenci, and M. Seri, Chaos 20 (2010) 023115.*

[CM] *N. Chernov e R. Markarian. Chaotic Billiards. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2006.*

[Co]*J.-P. Conze, Sur un critère de récurrence en dimension 2 pour les marches stationnaires, applications, Ergodic Theory Dynam. Systems 19 (1999), no. 5, 1233–1245.*

[D] *Donnay, Victor J. Using integrability to produce chaos: billiards with positive entropy. Comm. Math. Phys. 141 (1991), no. 2, 225–257. https://projecteuclid.org/euclid.cmp/1104248299.*

[DM] *Del Magno, G., Markarian, R.: A local ergodic theorem for non-uniformly hyperbolic symplectic maps with singularities. Ergod. Theory Dyn. Syst. 33, 83–107 (2013).*

[H] *Hartman, Philip (August 1960). "A lemma in the theory of structural stability of differential equations". Proc. A.M.S. 11 (4): 610–620. doi:10.2307/2034720. JSTOR 2034720.*

[KH] *Katok, A., Hasselblatt, B. (1995). Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809187.*

[KS] *A. Katok and J.-M. Strelcyn (in collaboration with F. Ledrappier and F. Przytycki), Invariant manifolds, entropy and billiards; smooth maps with singularities, LNM 1222, Springer-Verlag, Berlin-New York, 1986.*

[L1] *M. Lenci, Semi-dispersing billiards with an infinite cusp I, Comm. Math. Phys. 230 (2002), no. 1, 133–180.*

[L2] *M. Lenci, 2013 Exactness, K-property and infinite mixing Publ. Mat. Urug. 14 159–70.*

[L3] *Lenci, M. (2003). Aperiodic Lorentz gas: Recurrence and ergodicity. Ergodic Theory and Dynamical Systems, 23(3), 869-883. doi:10.1017/S0143385702001529.*

[L4] *Lenci, M.: Typicality of recurrence for lorentz gases. Erg. Th. Dynam. Systs. 26(3), 799–820 (2006).*

[Lo] *H. A. Lorentz, The motion of electrons in metallic bodies I, II, and III, Koninklijke Akademie van Wetenschappen te Amsterdam, Section of Sciences, 7 (1905), 438–453, 585– 593, 684–691.*

[LW] *C. LIVERANI and M. WOJTKOWSKI, Ergodicity in Hamiltonian systems, Dynamics Reported 4 (1995), 130-203, Springer-Verlag, New York.*

[P] *Pollicott, M. (1993). Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds (London Mathematical Society Lecture Note Series). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511752537*

[Pes1] *Ya. B. Pesin, "Characteristic Lyapunov exponents and smooth ergodic theory", Uspekhi Mat. Nauk, 32:4(196) (1977), 55–112; Russian Math. Surveys, 32:4 (1977), 55–114 .*

[Pes2] *Pesin, Y. (1992). Dynamical systems with generalized hyperbolic attractors: Hyperbolic, ergodic and topological properties. Ergodic Theory and Dynamical Systems, 12(1), 123-151. doi:10.1017/S0143385700006635.*

[R1] *V.A. Rokhlin, On the Fundamental Ideas of Measure Theory. Mat. Sb. (N.S.), 1949, 25(67):1, 107–150 (Russian). Translations Amer. Math. Soc., Series 1, 10 (1962), pp. 1–54.*

[R2] *V.A. Rokhlin, Lectures on the Entropy Theory of Measure-Preserving Transformations. Russ. Math. Surv. Vol. 22 (1967), No. 5, pp. 3–56. MR0217258.*

[S] *Ya. G. Sinai, "Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards", Uspekhi Mat. Nauk, 25:2(152) (1970), 141–192; Russian Math. Surveys, 25:2 (1970), 137–189.*

[Sc] *R. L. Schilling, Measures, Integrals and Martingales (Cambridge University Press, Cambridge, UK, 2005).*

[Sch]*K. Schmidt, On joint recurrence, Comptes Rendu Acad. Sci. Paris Sér. I Math. 327 (1998), no. 9, 837–842.*

[SLEC] *M. Seri, M. Lenci, M. Degli Esposti, and G. Cristadoro, J. Stat. Phys. 144 (2011) 124.*

[V] *M. Viana, Disintegration into conditional measure: Rokhlin's theorem. http://w3.impa.br/∼viana/out/rokhlin.pdf.*

[W] *P. Walters, An Introduction to Ergodic Theory, Springer, New York (1982).*

[W1] *Wojtkowski, M. (1985). Invariant families of cones and Lyapunov exponents. Ergodic Theory and Dynamical Systems, 5(1), 145-161. doi:10.1017/S0143385700002807.*

[W2] *Wojtkowski, Maciej 1986. Principles for the design of billiards with nonvanishing Lyapunov exponents. Communications in Mathematical Physics, Vol. 105, Issue. 3, p. 391.*