Annular billiards and their generalizations

Bachelor Thesis

Gergő Dénes

Supervisor: Dr. Péter Bálint

Budapest, Hungary

2023

Szakdolgozat-választás

A hallgató neve: Dénes Gergő	specializációja: Fizika BSc - fizikus
A záróvizsgát szervező tanszék neve: Elméleti Fizika Tanszék	

A témavezető neve: Bálint Péter - munkahelye: Matematika Intézet, Sztochasztika Tanszék - beosztása: egyetemi docens - email címe: balint.peter@ttk.bme.hu	A konzulens neve: Szunyogh László - tanszéke: Elméleti Fizika Tanszék - beosztása: egyetemi tanár - email címe: szunyogh.laszlo@ttk.bme.hu

A kidolgozandó feladat címe: Annulus biliárdok és általánosításaik

A téma rövid leírása, a megoldandó legfontosabb feladatok felsorolása:

Az annulus biliárdot úgy kapjuk, ha egy kör alakú tartományból kivágunk egy kör alakú ütközőt, és a fennmaradó részen tanulmányozzuk egy klasszikus tömegpont mozgását, mely a tartomány belsejében egyenes vonalú egyenletes mozgást végez, a határról pedig a geometriai optika szabályai szerint visszaverődik. Könnyen látható, hogy amennyiben a két kör koncentrikus, a mozgás integrálható. Más esetekben azonban a geometriai paraméterektől függően változatos dinamikai viselkedés figyelhető meg.

A szakdolgozó feladata, hogy feltárja a témara vonatkozó szakirodalmat, valamint vizsgáljon további általánosítási lehetőségeket.

A záróvizsga kijelölt tételei:

Dátum:

Hallgató aláírása:	Témavezető aláírása*:	Tanszéki konzulens aláírása:	A témakiírását jóváhagyom (tanszékvezető aláírása):

*A témavezető jelen feladatkiírás aláírásával tudomásul veszi, hogy a BME TVSZ 145. és 146.§ alapján az egyetem a képzési célok megvalósulása érdekében a szakdolgozatok, illetve diplomamunkák nyilvánosságát tartja elsődlegesnek. A hozzáférés korlátozása csak kivételes esetben, a dékán előzetes hozzájárulásával lehetséges.

Nyilatkozat

Alulírott **Dénes Gergő** a Budapesti Műszaki és Gazdaságtudományi Egyetem fizika BSc szakos hallgatója kijelentem, hogy ezt a szakdolgozatot meg nem engedett segédeszközök nélkül, önállóan, a témavezető irányításával készítettem, és csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból vettem, a forrás megadásával jelöltem.

Budapest, 2023. június 8.

_____

Dénes Gergő

Nyilatkozat

Alulírott **Dénes Gergő** a Budapesti Műszaki és Gazdaságtudományi Egyetem fizika BSc szakos hallgatója kijelentem, hogy a dolgozat digitális és kinyomtatott példánya mindenben megegyezik.

Budapest, 2023. június 8.

Dénes Gergő

Abstract

In this work, we cover – without proofs – the mathematical basics needed to study chaos theory, including measure theory, ergodic theory, hyperbolicity (which is at the heart of chaos theory), then establish a general method for finding regions of hyperbolicity (cone technique).

Then we introduce the billiard problem, cover the basic concepts that occur. After this, we study dispersing billiards, discuss why a stadium billiard is hyperbolic, prove that a circular billiard is NOT hyperbolic, then demonstrate some interesting phenomena occuring in annular billiards that consist of two circles.

Using the toolset and knowledge developed in the preceeding chapters and sections, we later examine a billiard that is made of two stadia, and find parameters that guarantee hyperbolicity.

A large portion of this work was taken from or paraphrased with the help of [1], and some figures are also taken from there. Figures where it isn't stated explicitly whether they are taken from another source or not, were made by the author (me, Gergő Dénes) using the TikZ package in LaTeX.

Contents

Chapter 1

Intruduction

In literature, chaos does not have a universal definition. However, a common definition for a chaotic dynamical system is that it satisfies the following properties:

(i) it is sensitive to initial conditions (hyperbolicity)

(ii) it is topologically transitive (mixing)

(iii) it has dense periodic orbits.

Property (iii) is easy enough to grasp and see whether it holds in a dynamical system, but properties (i) and (ii) are not so evident. In this chapter, we will introduce the most basic concepts needed to mathematically describe chaos through (i) and (ii). Furthermore, let us note that the above three properties characterizing chaos are understood in a purely topological perspective. In this work we include the additional viewpoint of ergodic theory with the presence of a natural invariant measure. This way we interpret properties (i) through the Oseledets theorem, while in the context of (ii), we interpret ergodicity/mixing with respect to the invariant measure.

We follow the most important definitions and derivations in the appendices of [1] for measure theory and ergodic theory, then for introducing hyperbolicity, we closely follow Chapter 3 in that very same book.

1.1 Measure theory basics

Definition 1.1.1 *Let X be a set. A nonempty set $\mathcal{F}$ constructed of the subsets of X is a* ***σ-algebra*** *if*

(i) if $A_i \in \mathcal{F}$, then $\forall i \geq 1 \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (it is closed under countable unions)

(ii) if $A \in \mathcal{F}$, then $A^C = X \backslash A \in \mathcal{F}$ (it is closed under taking complements)

$(X, \mathcal{F})$ together are called a ***measurable space****, the sets $A \in \mathcal{F}$ are called* ***measurable sets****.*

Any σ-algebra is closed under all countable combinations of unions, intersections, and all other elementary set-theoretic operations. Every σ-algebra contains X itself and $\varnothing$.

Definition 1.1.2 *For any family of σ-algebras $\{\mathcal{F}_i\}$, their intersection $\bigcap_i \mathcal{F}_i$ is also a σ-algebra. For any collection $\mathcal{C}$ of subsets of X let $\tilde{\mathcal{F}}(\mathcal{C})$ the intersection of all σ-algebras containing $\mathcal{C}$. Then $\tilde{\mathcal{F}}(\mathcal{C})$ is called the* ***σ-algebra generated by $\mathcal{C}$****.*

Definition 1.1.3 *If $\mathcal{C}$ denotes the collection of all open subsets of a topological space X, then $\tilde{\mathcal{F}}(\mathcal{C})$ is called the* ***Borel σ-algebra on X****.*

Definition 1.1.4 *A function $\mu : \mathcal{F} \to \mathbb{R} \cup \{+\infty\}$ is a* ***measure on*** *$(X, \mathcal{F})$ if*

(i) $\mu(A) \geq 0 \quad \forall A \in \mathcal{F}$ (μ is nonnegative)

(ii) $\mu(\varnothing) = 0$ (the empty set has zero measure)

(iii) If $\{A_{i=1}^{\infty}\} \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then $\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i)$ (μ is σ-additive)

Definition 1.1.5 *A* ***semi-algebra*** *is a nonempty collection $\mathcal{C}$ of subsets of X if*

(i) if $A, B \in \mathcal{C}$, then $A \cap B \in \mathcal{C}$ (closed under intersections)

(ii) if $A \in \mathcal{C}$, then $A^C = \bigcup_{i=1}^{n} A_i$ where $A_i \in \mathcal{C}$ and $i \neq j \implies A_i \cap A_j = \varnothing$

Theorem 1.1.1 ***Extension theorem*** *Let $\mathcal{C}$ be a semi-algebra and let $\mu : \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ such that*

(i) $X = \bigcup_{i=1}^{\infty} A_i$ *for some* $A_i \in \mathcal{C}$ *such that* $\mu(A_i) < \infty$

(ii) μ *is* σ*-additive (see Definition 1.1.4)*

Then there exists a unique measure $\tilde{\mu}$ *on* $\tilde{\mathcal{F}}(\mathcal{C})$ *that agrees with* μ *on* $\mathcal{C}$*.*

The proof of this theorem is omitted here.

An important application of this theorem is constructing the Lebesgue measure: take $\mathcal{C}$ as the collection of all open, closed and semi-open intervals in $\mathbb{R}$. Then $\mathcal{C}$ is a semi-algebra, and the length of an interval is a function on $\mathcal{C}$ that satisfies the above properties, thus there exists an **m** unique measure on the Borel σ-algebra on $\mathbb{R}$ that gives the length of intervals. This is called the **Lebesgue measure**.

Definition 1.1.6 μ *is a* ***finite measure*** *on* $(X, \mathcal{F})$ *if* $\mu(X) < \infty$*, and is a* ***probability measure*** *if* $\mu(X) = 1$*.*

Any Borel probability measure μ *on* $\mathbb{R}$ *can be defined by*

$$\mu((a, b]) = F(b) - F(a)$$

assuming the following hold:

(i) F *is nondecreasing*

(ii) F *is right-continuous*

(iii) $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to 0} F(x) = 0$*.*

Every function $F(x)$ with the above properties is a **distribution function** of a probability measure on $\mathbb{R}$. The restriction of **m** onto $[0, 1]$ is a probability measure as well.

Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be two measurable spaces with probability measures. A transformation $T : X \to Y$ is measurable if $\forall A \in \mathcal{G}$, we have $T^{-1}A \in \mathcal{F}$.

The image $T\mu$ of the measure μ is defined by $T\mu(A) = \mu(T^{-1}A) \quad \forall A \in \mathcal{G}$. An equivalent form of this fact can be written as:

$$\int_Y g d(T\mu) = \int_X g \circ T d\mu$$

for any bounded measurable function $g : Y \to \mathbb{R}$.

Notation 1.1.1 *For two measurable sets* $A, B \in \mathcal{F}$*, we write*

$$A = B \pmod 0 \iff \mu((A \backslash B) \cup (B \backslash A))$$

meaning that A *can be transformed into* B *by adding and/or removing some null sets.*

1.2 Ergodic theory basics

In the following, let $(X, \mathcal{F})$ be a measurable space.

Definition 1.2.1 *$T : X \to X$ is a **measurable transformation** if $T^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{F}$.*

From now on, let T be a measurable transformation.

Definition 1.2.2 *$T : X \to X$ is an **automorphism** if it is a bijection, and T^{-1} is also a measurable transformation.*

Note that the set $\{T^n\}, n \geq 0$ makes a semigroup, and if T is also an automorphism, then $\{T^n\}, n \in \mathbb{Z}$ makes a group. In both cases the group operation is the composition of functions.

For any point $x \in X$, the sequence $\{T^n x\}$ is called the orbit of x – for our purposes, we will refer to this as the **trajectory of** x.

Let $\mathcal{M}(X)$ be the set of all probability measures on $(X, \mathcal{F})$. $\forall \mu, \nu \in \mathcal{M}(X)$ and $0 < p < 1$ we get $p\mu + (1-p)\nu \in \mathcal{M}(X)$, meaning that $\mathcal{M}(X)$ is a **convex set**.

$T : X \to X$ induces a map $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ through $(T\mu)(A) = \mu(T^{-1}A) \quad \forall \mu \in \mathcal{M}(X), \forall A \in \mathcal{F}$.

Definition 1.2.3 *A measure μ is T-**invariant** if $T\mu = \mu$ (T preserves the measure μ).*

For an automorphism T, $T\mu = \mu \iff T^{-1}\mu = \mu$, so T and T^{-1} preserve the same measures. Let $\mathcal{M}_T(X)$ be the set of all T-invariant probability measures on X. Then $\mathcal{M}_T(X)$ is a convex subset of $\mathcal{M}(X)$.

A measure $\mu \in \mathcal{M}(X)$ is T-invariant if and only if for any measurable function $f : X \to \mathbb{R}$, we have

$$\int_X f \circ T d\mu = \int_X f d\mu$$

(if one of the integrals exist, the other exists as well).

For any $\mu \in \mathcal{M}(X)$ its image $\mu_1 = T\mu$ is given by

$$\int_X f \circ T d\mu = \int_X f d\mu_1$$

$T : X \to X$ induces a linear map U_T on the space of measurable functions $f : X \to \mathbb{R}$ by

$$(U_T f)(x) = (f \circ T)(x) = f(T(x))$$

Let $(X, \mathcal{F})$ be a measurable space and let $T : X \to X$ preserve a measure $\mu \in \mathcal{M}_T(X)$. $(X, \mathcal{F}, T, \mu)$ together are called a **measure-preserving transformation**.

Theorem 1.2.1 ***Poincaré recurrence theorem*** *Let $T : X \to X$ preserve a measure $\mu \in \mathcal{M}_T(X)$ and $\mu(A) > 0$ for some measurable set $A \subset X$. Then for μ-almost every point $x \in A$, we have*

$$T^{n_i}(x) \in A \quad \textit{for some sequence } n_1 < n_2 < \dots$$

Then,

$$T_A(x) = T^{n_A(x)}(x), \quad \textit{with} \quad n_A(x) = \min\{n \geq 1 : T^n(x) \in A\}$$

is defined almost everywhere on A, and is called the ***Poincaré return map****. This map preserves the conditional measure μ_A on A defined by $\mu_A(B) = \mu(A \cap B) \backslash \mu(A)$.*

The proof of this theorem is omitted.

This theorem is remarkable, as in layman terms, it states that nondissipative, deterministic dynamical systems will eventually return to their initial configurations.

Definition 1.2.4 *A measurable* ***set*** *$A \subset X$ is T-**invariant** if $T^{-1}(A) = A$.*

This happens if $T(A) \subset A$ and $T(A^C) \subset A^C$.

If T preserves a measure μ, then a measurable set A is T-invariant (mod 0) if $A = T^{-1}A$ (mod 0). If B is T-invariant (mod 0), then there exists a T-invariant set $\tilde{B}$ such that $B = \tilde{B}$ (mod 0).

Definition 1.2.5 *A* ***function*** *$f : X \to \mathbb{R}$ is T-**invariant** if $f = f \circ T$ (f is constant on every trajectory of the map T).*

If T preserves a measure μ, then $f : X \to \mathbb{R}$ is T-invariant (mod 0) if $f(x) = f(Tx)$ for μ-almost every point $x \in X$. In that case, there exists a T-invariant function $\tilde{f}$ such that $f = \tilde{f}$ (mod 0).

Definition 1.2.6 *A T-invariant measure $\mu \in \mathcal{M}_T(X)$ is* ***ergodic*** *if for any T-invariant set $A \subset X$ we have $\mu(A) = 0$ or $\mu(A) = 1$.*

A T-invariant measure μ is ergodic if and only if any T-invariant function $f : X \to \mathbb{R}$ will have $\mu(\{x : f(x) = c\}) = 1$ (f is constant almost everywhere).

If a measurable transformation $T : X \to X$ has a unique invariant measure μ, it will automatically be ergodic. In this case, T is said to be uniquely ergodic.

An example of a uniquely ergodic measure is rotating a circle by an irrational φ angle. In this case, only the Lebesgue measure is preserved. Hyperbolic systems (e.x. hyperbolic billiards) are typically not uniquely ergodic, as we can assign an invariant measure to every periodic trajectory.

Definition 1.2.7 ***Forward/backward time average:*** *Let $(X, \mathcal{F}, T, \mu)$ be a measure-preserving transformation and $f : X \to X$ a measurable function. Then $\forall x \in X$ the sequence $\{f(T^n x)\}$ is regarded as the value of f at time n. Then*

$$f_{\pm}(x) := \lim_{n \to \pm\infty} \frac{1}{n}\Big(f(x) + f(T^{\pm 1}x) + \cdots + f(T^{\pm(n-1)}x)\Big)$$

is called the forward/backward time average of the function f along the orbit of x (if the limit exists).

Theorem 1.2.2 ***Birkhoff Ergodic Theorem*** *Let $(X, \mathcal{F}, T, \mu)$ be a measure-preserving transformation and $f \in L^1_\mu(X) = \Big\{f : \int_X |f|^1 d\mu < \infty\Big\}$. Then*

(i) for almost every point $x \in X$ the limit $f_+(x)$ in Definition 1.2.7 exists.

(ii) if the function $f_+(x)$ exists, then $f_+(T^n x)$ exists for all n and $f_+(T^n x) = f(x)$ ($f(x)$ is T-invariant).

(iii) f_+ is integrable and $\int_X f_+ d\mu = \int_X f d\mu$

(iv) if μ is ergodic, then $f_+(x)$ is almost everywhere constant and its value is $\int_X f d\mu$.

The proof of this theorem is also omitted.

If $T : X \to X$ is an automorphism, then for almost every point $x \in X$ the limit $f_-(x)$ exists as well, and $f_-(x) = f_+(x) \pmod 0$.

The integral $\int_X f d\mu$ is regarded as the space-average of the function f. (iv) asserts that if μ is ergodic, then the time averages are equal to the space averages.

This theorem is also remarkable, and we usually take result (iv) as granted in statistical physics.

Definition 1.2.8 *A map $T : X \to X$ preserving a measure μ is **mixing** if for all pairs of measurable subsets $A, B \subset X$*

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

As $\mu(A) = \mu(T^{-n}A)$, this is the same as

$$\lim_{n\to\infty} |\mu(T^{-n}A \cap B) - \mu(T^{-n}A)\mu(B)| = 0$$

meaning that the events $T^{-n}A$ and B become asymptotically independent as $n \to \infty$.

Note that $x \in T^{-n}A \iff T^n(x) \in A$, meaning that we are speaking about the events $x \in B$ and $T^n(x) \in A$ (these quantities characterize x at time 0, and x at time n). Mixing is commonly interpreted as asymptotic independence of the distant future from the present, as the system 'forgets where it started'.

Also note that mixing implies ergodicity: Let A be an invariant set and $B = A$. Then

$$\mu(T^{-n}A \cap A) = \mu(A \cap A) = \mu(A)$$

But from the mixing property, we get

$$\lim_{n\to\infty} \mu(T^{-n}A \cap A) = \mu(A)\mu(A)$$

giving $\mu(A) = \mu(A)^2$, meaning that $\mu(A)$ is either 0 or 1, which means that μ is ergodic.

Definition 1.2.9 *Let $(X, \mathcal{F})$ be a measurable space. A dynamical system with continuous time, also known as a **flow** is a one-parameter family of measurable transformation $\{S^t\} : S^t : X \to X, \quad t \in \mathbb{R}$ such that $S^{t+s} = S^t \circ S^s$ and the map $X \times \mathbb{R} \to X$ defined by $(x, t) \mapsto S^t x$ is measurable.*

For every point $x \in X$, the set $\{S^t x\} \quad t \in \mathbb{R}$ is called the orbit, or in our case more relevantly the trajectory of x. In billiards, X is a topological space and $\{S^t x\}$ is a continuous curve $\forall x \in X$.

$\{S^t\}$ preserves a measure $\mu \in \mathcal{M}(X)$ if $\mu(S^t A) = \mu(A)$ for all measurable subsets $A \subset X$ and all $t \in \mathbb{R}$.

Invariant measures, the Birkhoff Ergodic Theorem and mixing can all be extended to flows quite intuitively, so the specifics of these are omitted.

1.3 Lyapunov exponents and hyperbolicity

Let M be a compact Riemannian manifold, $N \subset M$ an open and dense subset and $F : N \to M$ a $C^r \quad r \geq 2$ diffeomorphism of N onto $F(N)$. Assume that F preserves a probability measure μ on M and $\mu(\tilde{N}) = 1$, where

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

is the set where every iteration of F is defined.

Theorem 1.3.1 ***Oseledets theorem****: Let* $\log^+ s = \max\{\log s, 0\}$ *and* D_xF *is the derivative map of* F *at point* x *and*

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

Then, there exists an F*-invariant set* $H \subset \tilde{N}, \mu(H) = 1$ *such that* $\forall x \in H$*, there exists a* DF*-invariant decomposition of the tangent space*

$$\mathcal{T}_xM = E_x^{(1)} \oplus \cdots \oplus E_x^{(m)}$$

with some $m = m(x)$ *such that for* $V_x^i = E_x^{(1)} \oplus \cdots \oplus E_x^{(i)} \quad (0 < i \leq m)$ *we have that* $\forall v(\neq 0) \in V_x^{(i)}$

$$\lim_{n \to \pm\infty} \frac{1}{n} \log ||D_xF^n v|| = \lambda_x^{(i)} \tag{1.1}$$

where $\lambda_x^{(1)} < \cdots < \lambda_x^{(m)}$.

The proof of this theorem is also omitted.

Definition 1.3.1 *The* $\lambda_x^{(i)}$ *values in the above theorem are called* ***Lyapunov exponents*** *of* F *at* x *and have* $k_i = \dim E_x^{(i)}$ ***multiplicities****.*

Lyapunov exponents may be defined regardless of invariant measures, as long as DF-invariant decomposition and the limits (1.1) exist at $x \in \tilde{N}$.

Lyapunov exponents and their multiplicities are invariant under F. If F is ergodic, then the Lyapunov exponents and their multiplicities are almost everywhere constant.

(1.1) can be restated in an equivalent way: Let $\lambda_x = \min_i |\lambda_x^{(i)}|$. Then $\forall \varepsilon > 0$ $\exists C(x, \varepsilon) > 0$ such that

$$||D_xF^{-n}(v)|| \leq C(x, \varepsilon)e^{-n(\lambda_x - \varepsilon)}||v|| \quad \forall v \in E_x^u \tag{1.2}$$

and

$$||D_xF^n(v)|| \leq C(x,\varepsilon)e^{-n(\lambda_x-\varepsilon)}||v|| \quad \forall v \in E^s_x \tag{1.3}$$

If $\lambda^{(i)}_x > 0$, then nonzero tangent vectors $v \in E^{(i)}_x$ grow exponentially in the future and contract exponentially in the past. The opposite is true for $\lambda^{(i)}_x < 0$. If $\lambda^{(i)}_x = 0$, then the corresponding tangent vectors do not expand or contract at an exponential rate, but they might still grow or shrink linearly or polynomially in n.

In physics, for non-dissipative systems we get that Lyapunov exponents add to 0, meaning that if there exists a positive exponent, at least one negative exponent also exists, which corresponds to stretching in one direction and contraction in another in the tangent space. This is also going to be the case for billiards.

Definition 1.3.2 *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 $x \in M$, the tangent space is given by $\mathcal{T}_xM = E^u_x \oplus E^s_x$, where

$$E^u_x = \bigoplus_{\lambda^{(i)}_x > 0} E^{(i)}_x \qquad \text{and} \qquad E^s_x = \bigoplus_{\lambda^{(i)}_x < 0} E^{(i)}_x$$

corresponding to the unstable and stable subspaces of F. The words 'unstable' and 'stable' indicate that if any tangent vector in E^u_x gets transformed into E^u_x by D_xF, its length will increase, leading to trajectories close to each other getting further apart, while tangent vectors in E^s_x get transformed into E^s_x and have shorter length, leading to trajectories close to each other to get even closer.

Definition 1.3.3 *F is a **hyperbolic map** if μ-almost every point $x \in M$ is hyperbolic.*

If F is a hyperbolic map, then $\lambda_x > 0$ and $C(x,\varepsilon) > 0$ (in the sense of (1.2) and (1.3)) are measurable functions on M.

Definition 1.3.4 *Hyperbolicity is **uniform** if λ_x and $C(x,\varepsilon)$ can be made constant, meaning $\exists \lambda, C > 0 : \quad \forall x \in M \wedge n \geq 1$:*

$$||D_xF^{-n}(v)|| \leq Ce^{-n\lambda}||v|| \quad \forall v \in E^u_x$$

$$||D_xF^{n}(v)|| \leq Ce^{-n\lambda}||v|| \quad \forall v \in E^s_x$$

We will see that dispersing billiards are uniformly hyperbolic, stadia are non-uniformly hyperbolic (this latter will be a result of circular billiards not being hyperbolic at all).

1.4 Cone technique for proving hyperbolicity

The method described below is quite useful, and will be applied in detail in Chapter 4.

A general method for establishing hyperbolicity for $x \in M$ and constructing their stable and unstable subspaces E^s_x and E^u_x are as follows (for sake of simplicity and our problem, we take $\dim M = 2$, but it can be extended into higher dimensions):

Let $x_n = F^n(x) \quad \forall n \in \mathbb{Z}$, and choose some nonzero vectors in the tangent space of M at point x_{-n} and $n \geq 0$, so $v_{-n} \in \mathcal{T}_{x_{-n}}M$. Then, the following equation holds:

$$\lim_{n\to\infty} \operatorname{span}\{D_{x_{-n}}F^n(v_{-n})\} = E^u_x \tag{1.4}$$

The above equation states that if we iterate the derivative map on a certain v_{-n} vector in the tangent space 'large enough' n times, the resulting tangent vector will be in the unstable space, as its component in the stable subspace gets transformed into 0, and its component in the unstable subspace gets larger and larger.

(1.4) does not hold for all choices of the set $\{v_{-n}\}$, but for some typical choices it does. (1.1) suggests that if $\lambda_x > 0$ and $v_{-n} = c_{-n,u}v_{-n,u} + c_{-n,s}v_{-n,s}$ with $v_{-n,u} \in E^u_{x_{-n}}, v_{-n,s} \in E^s_{x_{-n}}$, then

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

with $w_{-n,u} \in E^u_x$ and $w_{-n,s} \in E^s_x$ some unit vectors.

If $0 < c_{\min} \leq |c_{-n,u}|, \quad |c_{-n,s}| \leq c_{\max} < \infty$ (the coefficients are bounded away from zero and infinity), then (1.4) holds. As long as $\dfrac{|c_{-n,s}|}{|c_{-n,u}|}$ grows slower than $e^{2\lambda_x n}$ as $n \to \infty$, (1.4) still holds, so what we need to do is choose initial vectors v_{-n} not too close to the stable spaces $E^s_{x_{-n}}$. For choosing v_{-n}, the method of cones can be introduced. For the next definition, it is again important to note that we work in 2 dimensions, but this method can be extended into higher dimensions.

Definition 1.4.1 *Let $L \subset \mathcal{T}_xM$ be a line and $\alpha \in \left(0, \dfrac{\pi}{2}\right)$. A **cone** C with axis L and opening α is the set of all tangent vectors $v \in \mathcal{T}_xM$ that make angle $\nu < \alpha$ with the line L.*

∂C is made up of vectors making angle α with L (the zero vector satisfies this as well, so it belongs to the boundary).

Assuming that almost every point $x \in M$ is hyperbolic and we have a cone field – i.e. a cone defined for every point $x \in M$ that depends continuously on x –, then $C^u_x \subset \mathcal{T}_xM$

satisfies $D_xF(C_x^u) \subset C_{Fx}^u$ (the cone is invariant) at almost all points $x \in M$ (with L_x and α_x measurable functions of x). Then it can be shown that

$$E_x^u = \bigcap_{n\geq 0} D_{x_{-n}}F^n(C_{x_{-n}}^u)$$

Using this, $\forall v_{-n} \in C_{x_{-n}}^u$ initial vectors satisfy (1.4).

Similarly, one can construct the stable subspace E_x^s to get $E_x^s = \bigcap_{n\geq 0} D_{x_n}F^{-n}(C_{x_n}^s)$ with an appropriately defined C_x^s.

The above mean that if we find invariant cones in a hyperbolic system, the unstable subspace is going to be the infinite intersection of the transformed invariant unstable cones and the stable subspace is the infinite intersection of the images of the stable cones by the inverse of the transformation. However, if the system is not hyperbolic (meaning that the stable and unstable subspaces are empty), these cones can still exist (we will see an example of this later with circular billiards).

M. Wojtkowski's results [2] imply that although invariance is not enough, **strict invariance** of the cones imply hyperbolicity for maps that preserve an absolutely continuous measure (in dimensions greater than 2, the result is true for 'symplectic' maps instead, but that is out of the scope of our work). By strict invariance we mean

$$D_xF(C_x^u) \subset \text{int}C_{Fx}^u \cup \{0\} \tag{1.5}$$

$$D_xF^{-1}(C_x^s) \subset \text{int}C_{F^{-1}x}^s \cup \{0\} \tag{1.6}$$

where C_x^u and C_x^s are the unstable- and stable cones respectively.

Note that the conditions we derived hold only for uniformly hyperbolic maps. For nonuniformly hyperbolic maps, the cone method can be applied with the addition that there exist a measurable function $n_x : M \to \mathbb{N}$ such that for all $n \geq n_x$ the derived results hold (this can be referred to as 'eventual invariance' of the cones).

Chapter 2

The billiard problem

Now that we have laid the basics of the mathematics needed for our problem, we introduce the billiard problem and Jacobi coordinates, according to Chapters 2 and 3 in [1].

2.1 What is a billiard?

Let $D_0 \subset \mathbb{R}^2$ be a bounded open connected domain and $D = \overline{D_0}$. We will assume that the boundary ∂D is a finite union of at least C^3 smooth compact curves.

$$\partial D = \Gamma = \Gamma_1 \cup \cdots \cup \Gamma_r$$

where each Γ_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 a_i and b_i, assuming $\ell \geq 3$.

A D domain described above is called a **billiard table** and $\Gamma_1, \ldots, \Gamma_r$ the **walls** of the billiard table, or the components of ∂D.

If Γ_i is not a closed curve ($f(a_i) = f(b_i)$), we call it an arc.

From now on, we also assume that the boundary components Γ_i intersect each other only at their endpoints (note that at these points, ∂D might not be 1-smooth – these are called corners –, but this will not be a problem).

Let us parametrize each Γ_i with its arclength in such a way that int$D$, the interior of the billiard table lies to the left of Γ_i. Then, because of the parametrization, the tangent vectors become unit vectors: $|f_i'| = 1$.

From now on, we also assume that on every Γ_i, the second derivative, f_i'' is either identically 0, or is never 0. This means that we disect the boundary to 'completely flat' walls and curves without inflection points.

Based on their second derivative f'' (these are vectors), we classify three types of curves:

- **Flat walls:** $f'' = 0$
- **Focusing walls:** $f'' \neq 0$ and points inside D
- **Dispersing walls:** $f'' \neq 0$ and points outside D

[1] – the basis of this chapter – has a marvelous figure for demonstrating how a 'general' billiard table looks like, which we include below:

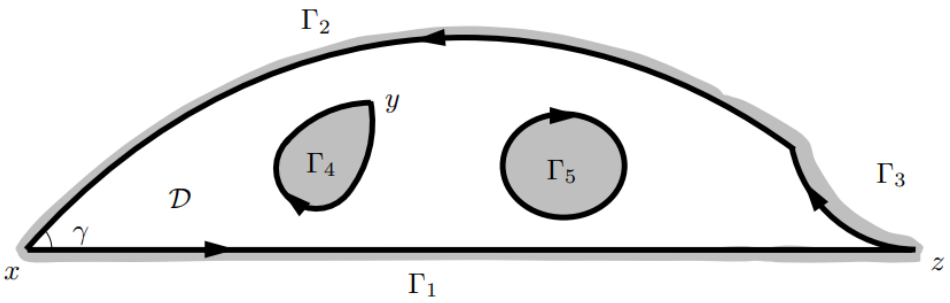

Figure 2.1: A billiard table $D = \mathcal{D}$ with different types of walls: Γ_1 is flat, Γ_2 is focusing, Γ_3 is dispersing, Γ_4 is a closed curve with a corner point, Γ_5 is a smooth closed curve. Corner x has a positive interior angle γ, corner z is a 'cusp' (the derivatives of Γ_3 and Γ_1 point in the same direction at z). The orientation of the curves are shown by arrows, and the unreachable parts of $\mathbb{R}^2$ are colored with grey (not just the grey area, but anyhting outside $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ also belongs to this set). Figure 2.2 in [1]

We can also define the curvature on each Γ_i, with $\kappa = 0$ for flat components, $\kappa = -||f''||$ for focusing components and $\kappa = ||f''||$ for dispersing components. Note that the curvature κ is a function of the base point on a given Γ_i boundary component, it is generally not constant.

The above given construction for billiards suggests an analogy with geometric optics: the particles that trace out trajectories in the system can be interpreted as light rays, while walls can be viewed as perfect mirrors. The terms 'focusing' and 'dispersive' are also quite intuitive, as light rays hitting a focusing mirror converge and light rays hitting a dispersing mirror diverge from each other.

Note that based on the above analogy with geometric optics, corners might give rise to interesting phenomena in billiards – we omit their investigation completely in this work.

Let $q \in D$ denote the position of a particle, and $v \in \mathbb{R}^2$ its velocity vector. Both of these quantities are functions of $t \in \mathbb{R}$ time. When the particle is moving freely inside the table, we assume no dissipation, so

$$\ddot{q} = \dot{v} = 0$$

When the particle collides with a smooth part of the boundary at q_c, its velocity vector gets reflected across the tangent line to that part of the boundary.

If a moving particle were to hit a corner, its trajectory would not be defined from then on. This is one of the reasons we omit the investigation of these points.

In a grazing collision, v^- is tangent to a smooth, dispersing part of Γ.

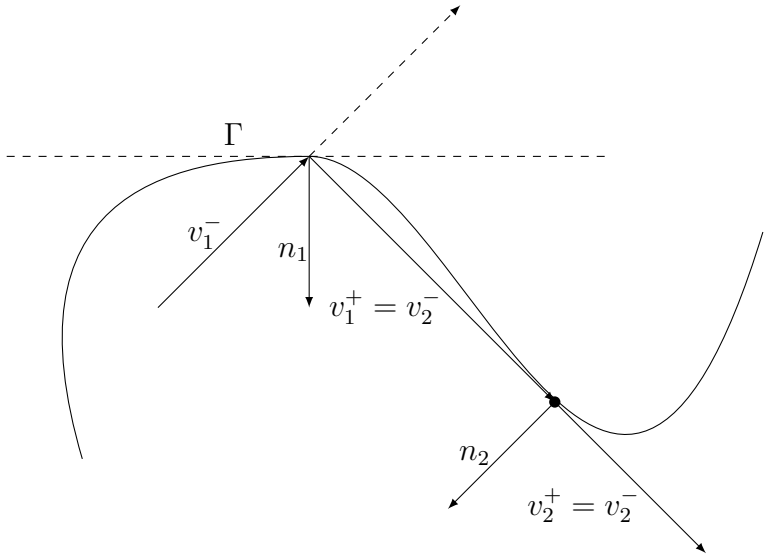

Figure 2.2: A particle bouncing off of a focusing boundary, then suffering a grazing collision

Because $|v|$ is unchanging, it is convenient to choose a coordinate system in which v's length is set to 1, meaning that the particle's velocity is a unit vector and the set of all possible v-s trace out a circle, so $v \in S^1$. The state of the particle at any time is specified by q and v, so the phase space of the system is given by

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

Let $\tilde{\Omega} \subset \Omega$ denote the set of states (q, v) in which the trajectory of the particle is defined for all $t \in \mathbb{R}$. Thus we obtain the flow of the particle:

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

$\forall t \in \mathbb{R}$ with the note that before and after a collision, (q, v^-) and (q, v^+) are to be identified with each other in the set of states (this is a convention). Also note that $\Phi^0 = \mathrm{Id}$ and $\Phi^{t+s} = \Phi^t \circ \Phi^s$, and everything discussed at Definition 1.2.9 holds.

2.2 Coordinates of the flow

The flow can be described in coordinates (x, y, ω) with $(x, y) = q$ being the usual carthesian coordinates and $\omega \in [0, 2\pi)$ the counterclockwise angle between v and the x-axis.

For flow without collisions:

$$\Phi^t : \Omega \to \Omega : (x^-, y^-, \omega^-) \mapsto (x^+, y^+, \omega^+) = (x^- + t\cos\omega, y^- + t\sin\omega, \omega^-)$$

For describing a collision, let $(\bar{x}, \bar{y}) \in \Gamma$ denote a collision point with $\mathbf{h}$ tangent vector to Γ and γ angle between $\mathbf{h}$ and the x-axis. Below we compute the post-collison coordinates of the phase point $X_+ = (x_+, y_+, \omega_+)$ which depend on the pre-collision phase point $X_- = (x_-, y_-, \omega_-)$. Let s^- denote the time before the collision and $s^+ = s^- + t$ the time elapsed since, meaning $X_+ = \Phi^t X_- = \Phi^{s_+ - s_-} X_-$. Let ψ denote the angle between v^+ and $\mathbf{h}$ (See Figure 2.3).

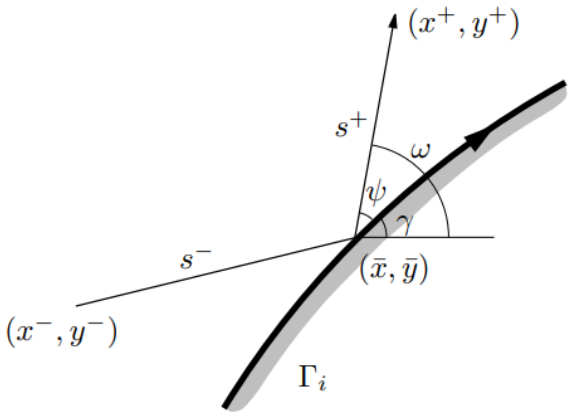

Figure 2.3: A collision on a curve Γ_i, with noted parameters that help describe the flow at the point of collision. It is not noted on the figure, but γ is the angle between the tangent vector at $(\bar{x}, \bar{y})$, $\mathbf{h}$ and the x-axis. Figure 2.8. in [1]

Then

$$\begin{aligned} x^- &= \bar{x} - s^- \cos\omega^-, & x^+ &= \bar{x} + s^+ \cos\omega^+, \\ y^- &= \bar{y} - s^- \sin\omega^-, & y^+ &= \bar{y} + s^+ \sin\omega^+, \end{aligned} \tag{2.1}$$

$$\omega^- = \gamma - \psi, \qquad \omega^+ = \gamma + \psi.$$

Now we aim to compute the differentials of the above quantities. Let r denote the arclength parameter on Γ, then

$$d\bar{x} = \cos\gamma dr, \qquad d\bar{y} = \sin\gamma dr, \qquad d\gamma = -\kappa dr \tag{2.2}$$

where κ is the curvature of the boundary component at $(\bar{x}, \bar{y})$.

Taking differentials of all 6 equations in (2.1), we get the following quantities:

$$\begin{aligned} dx^+ &= \cos\gamma dr + \cos\omega^+ ds^+ - s^+ \sin\omega^+ d\omega^+ \\ dy^+ &= \sin\gamma dr + \sin\omega^+ ds^+ + s^+ \cos\omega^+ d\omega^+ \\ d\omega^+ &= -\kappa dr + d\psi \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} dx^- &= \cos\gamma dr - \cos\omega^- ds^- + s^- \sin\omega^- d\omega^- \\ dy^- &= \sin\gamma dr - \sin\omega^- ds^- - s^- \cos\omega^- d\omega^- \\ d\omega^- &= -\kappa dr - d\psi \end{aligned} \tag{2.4}$$

It can be shown that the differential form $dx \wedge dy \wedge d\omega$ is preserved by Φ^t, so the Lebesgue measure $dxdyd\omega$ is also preserved on Ω.

2.3 Jacobi coordinates and wavefront curvature

Consider a point $X = (x, y, \omega) \in \Omega$ and a tangent vector in the tangent space of Ω at X: $dX = (dx, dy, d\omega) \in \mathcal{T}_X\Omega$. A more convenient coordinate system can be defined by $(d\eta, d\xi, d\omega)$, where

$$d\eta = \cos\omega dx + \sin\omega dy, \quad \text{and} \quad d\xi = -\sin\omega dx + \cos\omega dy \tag{2.5}$$

These are called **Jacobi coordinates**. $d\eta$ is the component of the vector (dx, dy) along v, and $d\xi$ is the orthogonal component of this vector. It can be seen that if there are no collisions between X and $\Phi^t X$, then

$$d\xi_t = d\xi + t d\omega, \qquad d\omega_t = d\omega, \qquad d\eta_t = d\eta \tag{2.6}$$

Now we consider what happens to the quantities at a collision. Denote by $d\xi^-, d\omega^-$ the precollisional-, $d\xi^+, d\omega^+$ the postcollisional values of the quantities $d\xi, d\omega$. (2.3) and (2.4) imply

$$d\xi^+ = -d\xi^- \tag{2.7}$$

and

$$d\omega^+ = -\frac{2\kappa}{\cos\varphi} d\xi^- - d\omega^- \tag{2.8}$$
$$d\eta^+ = d\eta^-$$

During the flow $d\eta$ is constant, and because neither $d\omega$ nor $d\xi$ depend on it, it is enough to only use the coordinates $(d\xi, d\omega)$ to describe the tangent space instead of $(d\eta, d\xi, d\omega)$ ($d\eta$ would not stay constant if there was dissipation or a potential).

These $(d\xi, d\omega)$ quantities correspond to lines in the tangent space $L \subset \mathcal{T}_x^{\perp}\Omega$. We have not yet explicitly stated, but by calculating the above quantities, we have been describing the action of the derivative $D\Phi^t$ of the flow Φ^t on these tangent lines.

The ratio of these quantities, $\dfrac{d\omega}{d\xi}$ define the slope of these lines in Jacobi coordinates. Let

$$B = \frac{d\omega}{d\xi} \tag{2.9}$$

with adding that if $d\xi = 0$ and $d\omega \neq 0$, we set $B = \infty$ (whether this is positive or negative ∞, it will not matter).

B, or put simply, the sign of B has a geometric meaning as well: $(d\xi, 0)$ corresponds to a displacement of $dq = (dx, dy)$ of $q = (x, y)$ in the direction perpendicular to v because of the definition of $d\xi$ in (2.3). $d\omega$ corresponds to a displacement dv of the vector v, so $B > 0$ if and only if dq and dv point in the same direction - see the figure below:

$B > 0$ $\qquad$ $B < 0$

Figure 2.4: Geometric meaning of the sign of B

Let $\gamma' \subset \Omega$ be a curve pasing through a point $X = (q, v)$ and tangent to a line $L \subset \mathcal{T}_x^{\perp}\Omega$. Assuming $B \neq \infty$, the projection $\sigma' = \pi_q(\gamma')$ of the curve γ' onto the table D

is a curve which is orthogonal to v at the point q. (π_q and π_v are the natural projections of Ω onto D and S^1 such that $\pi_q(q', v') = q'$ and $\pi_v(q', v') = v'$.)

Every point $X' = (q', v') \in \gamma'$ moves along a trajectory, $(q' + sv', v')$, which traces a segment of the line $q' + sv'$ on the billiard table (for small enough s that $q' + sv'$ stays inside D). Then, $A = \{q' + sv' : (q', v') \in \gamma'\}$ is a C^1 smooth family of directed lines in D. A is a 2-dimensional family of lines, as $v' = v'(q')$ is a function of $q' \in \sigma'$.

Let σ be the orthogonal cross-section of this family passing through q (this curve is perpendicular to every line in A). If we equip every point $q'' \in \sigma$ with a unit normal vector v'' to σ pointing in the direction of motion, we get a curve $\gamma \subset \Omega$.

This γ contains $X = (q, v)$ and it is tangent to γ' and L at X.

Definition 2.3.1 *A smooth curve $\sigma \subset D$ equipped with a continuous family of unit normal vectors described above is called a* ***wavefront****.*

The sign of B corresponds to different types of wavefronts: $B > 0$ means the wavefront is **dispersing**, $B < 0$ means the front is **focusing**, $B = 0$ means the front is **flat**, and $B = \infty$ means the front is degenerate - it is completely focused in a **focusing point** (also called degenerate). For clarity, [1] has a handy figure for this:

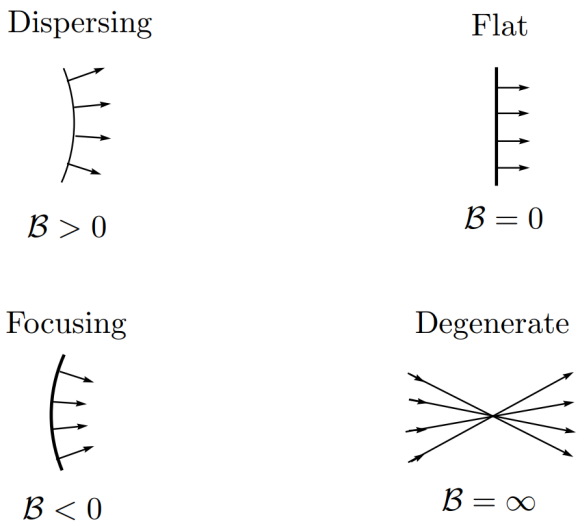

Figure 2.5: Types of wavefronts depending on the sign and finiteness of B (Symbol $\mathcal{B}$ is used instead of B). Figure 3.8. in [1]

B can be identified as the curvature of the wavefront. In the following it will become clear why:

From (2.6), it can be seen that for flow without collision, starting with B, after time t, we will get

$$B_t = \frac{d\omega}{d\xi} = \frac{B}{1+tB} = \frac{1}{t+\frac{1}{B}} \tag{2.10}$$

This equation makes it evident why B is associated with the wavefront's curvature. Assume $B_0 = \infty$, then through defocusing, the wavefront will become a circle with radius t, and B_t will be $\frac{1}{t}$, which is the curvature of this very circle.

For collisions, it is also clear from (2.7) and (2.8) that if $B = B^-$ before collision, then after the collision it will be

$$B^+ = B^- + \frac{2\kappa}{\cos\varphi} \tag{2.11}$$

For an n long series of these collisions, with $t_i - t_{i-1}$ distance between the i-th and $i-1$-st consecutive collisions, and the wall having κ_i curvature at the i-th collision, being hit by the trajectory at incident angle φ_i, we can collect the equations into a single continued fraction for the slope B_t:

$$B_t = \cfrac{1}{t - t_n + \cfrac{1}{\cfrac{2\kappa_n}{\cos\varphi_n} + \cfrac{1}{\cfrac{1}{t_n - t_{n-1} + \cfrac{1}{\cfrac{2\kappa_{n-1}}{\cos\varphi_{n-1}} + \cfrac{1}{\ddots + \cfrac{1}{t_1 + \frac{1}{B}}}}}}}} \tag{2.12}$$

If B is the curvature right before a collision, then $t_1 = 0$, the formula still holds.

2.4 The collision map

The problem can also be discretized to only map from collision to collision. It is easy to realize that $\Gamma \times S^1$ is a hypersurface describing the collisions at the boundary of the billiard table, but because we know the postcollisional vectors are related to the precollisional vectors through the normal vectors on Γ, it is enough to describe the space of postcollisional vectors.

For this, start by defining $\mathcal{M} = \bigcup_i \{x = (q, v) \in \Omega : \quad q \in \Gamma_i, \langle v, n\rangle \geq 0\}$ where n is the unit normal vector to the i-th boundary component Γ_i, pointing inside D ($\langle v, n\rangle$ is

the standard Euclidean inner product). $\mathcal{M}$ is a 2-dimensional submanifold in Ω, and it is called the **collision space**.

We can use the arclength on Γ for one of the coordinates parametrizing $\mathcal{M}$. On each Γ_i component, let $r \in [a_i, b_i]$ with of course $|\Gamma_i| = b_i - a_i$, and taking (a_i, b_i) disjoint intervals on $\mathbb{R}$ for different values of i. On smooth closed Γ_i, a_i can be identified with b_i, meaning that r is a cyclic parameter. For each point $x \in \mathcal{M}$, let $\varphi = \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$ be the angle between n and v with φ being positive for clockwise- and negative for counterclockwise oriented angles.

Every component of the collision space $\mathcal{M}$ has its each parametrization, but as long as the boundary components are connected, $\mathcal{M}$ can be made into a connected surface as well. If a boundary component is disjoint from some boundaries, then the collision space belonging to this obstacle will also be separate from those boundaries' collision space.

Let $x = (q, v) \in \mathcal{M}$. If q is not a grazing collision or at a corner, and $\langle v, n \rangle > 0$, then its trajectory $\Phi^t x$ is defined for some interval $0 < t < \varepsilon$. If the trajectory $\Phi^t x$ for $x \in \mathcal{M}$ is defined for some interval of time $(0, \varepsilon)$, then this trajectory can be continued up to the point x on the surface $\Gamma \times S^1$ and we denote the value of time assigned to this intersection by $\tau(x) > 0$, which we call the return time. Since $|v| = 1$, $\tau(x)$ is equal to the distance the trajectory covers from x before the next collision.

Let $\tilde{\Omega}_c \subset \tilde{\Omega} \subset \Omega$ denote the set of trajectories that have collisions in them.

Definition 2.4.1 *Then let $\tilde{\mathcal{M}} = \mathcal{M} \cap \tilde{\Omega}_c$, for which we can define the* ***collision map*** $\mathcal{F} : \tilde{\mathcal{M}} \to \tilde{\mathcal{M}} : \quad \mathcal{F}(x) = \Phi^{\tau(x)+}(x)$.

Let $x = (r, \varphi) \in \text{int}\mathcal{M}$, then $\mathcal{F}(x) = (r_1, \varphi_1) \in \text{int}\mathcal{M}$. Let $(\bar{x}, \bar{y}), (\bar{x_1}, \bar{y_1}) \in \partial D$ be the coordinates of the boundary points corresponding to r and r_1. Also let ω be the angle made by the trajectory between these two points and the x-axis. Then with $\tau = \tau(x)$,

$$\bar{x_1} - \bar{x} = \tau \cos\omega, \qquad \bar{y_1} - \bar{y} = \tau \sin\omega \tag{2.13}$$

Let γ, ψ be the same as in Figure 2.3 then equations (2.1) and (2.2) hold again. Similar notations for $(\gamma_1, \psi_1, \kappa_1)$ can be used at the point x_1. Because of the definition of ω, we get

$$\omega = \gamma + \psi = \gamma_1 - \psi_1$$

Taking differentials of this yields

$$d\omega = -\kappa dr + d\psi = -\kappa_1 dr_1 - d\psi_1 \tag{2.14}$$

Taking the differentials of (2.13):

$$\cos\gamma_1 dr_1 - \cos\gamma dr = \cos\omega d\tau - \tau\sin\omega d\omega$$

$$\sin\gamma_1 dr_1 - \sin\gamma dr = \sin\omega d\tau + \tau\cos\omega d\omega$$

From the above 4 equations, we get

$$\sin\psi_1 dr_1 + \sin\psi dr = \tau d\omega \tag{2.15}$$

Solving (2.14) and (2.15) for dr_1 and $d\psi_1$, and changing the parameter ψ to the angle of incidence φ through $\psi = \frac{\pi}{2} - \varphi$, we get

$$dr_1 = -\Big(\frac{\tau\kappa}{\cos\varphi_1} + \frac{\cos\varphi}{\cos\varphi_1}\Big)dr - \frac{\tau}{\cos\varphi_1}d\varphi \tag{2.16}$$

and

$$d\varphi_1 = -\Big(\frac{\tau\kappa\kappa_1}{\cos\varphi_1} + \kappa + \kappa_1\frac{\cos\varphi}{\cos\varphi_1}\Big)dr - \Big(\frac{\tau\kappa_1}{\cos\varphi_1} + 1\Big)d\varphi \tag{2.17}$$

Therefore, there exists a matrix $D\mathcal{F}$ which acts on the vector $(dr, d\varphi)$ at $x = (r, \varphi)$ and transforms it to $(dr_1, d\varphi_1)$ at $x_1 = (r_1, \varphi_1)$ through the parameters of the collisions and the distance between the collision points. This is called the **derivative map**.

Chapter 3

Dispersing billiards, stadia and annular billiards

We have now seen a way to describe billiard tables generally. In this chapter, we aim to explore different types of simple billiard tables. This chapter still heavily relies on [1]. For dispersing billiards, it will follow Chapter 4, and for stadium-shaped billiards, it will follow some of Chapter 8 in the book. In a later section we will mention phenomena appearing in [4], and we will share useful insight into what really causes hyperbolicity described in [3].

In Chapter 4, we will aim to use our knowledge gathered in this chapter to establish hyperbolicity on our own for a specific class of billiard tables.

3.1 Dispersing billiards

Definition 3.1.1 *A billiard table $D \subset \mathbb{R}^2$ is **dispersing** if all walls $\Gamma_i \subset \Gamma = \partial D$ are dispersing.*

In [1] there are further classifications of these billiards that include whether or not the billiard has corners (in that case, does it have cusps) and whether the billiard's horizon is bounded or not. For the notion of bounded horizon, see Definition 3.1.3 below.

First we have to define Tor^2:

Definition 3.1.2 *Let $\mathbb{K}_1 = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Then $\mathbb{R}^2$ can be covered by parallel translations of $\mathbb{K}_1$. Tor^2 is obtained by identifying the opposite sides of the square $\mathbb{K}_1$ ($(0, y) \equiv (1, y)$ and $(x, 0) \equiv (x, 1)$ in the values x, y allowed on $\mathbb{K}_1$).*

The standard projection of $\mathbb{R}^2$ onto $\mathbb{K}_1$ just takes the coordinates and produces the fractional parts of them: $(x, y) \mapsto (x - \lfloor x \rfloor, y - \lfloor y \rfloor)$, i.e. the standard projection transforms trajectories in $\mathbb{R}^2$ into directed straight lines on Tor^2. (For more information, see Chapter 1 in [1])

Definition 3.1.3 *A billiard table $D \subsetneq Tor^2$ has **bounded horizon** if $\tilde{\Omega}_f = \varnothing$, where $\tilde{\Omega}_f$ is the union of all collision-free trajectories. If this is untrue, the horizon is unbounded.*

In dispersing billiards, the curvature of ∂D is always positive, $\kappa > 0$.

Because of our choice in Chapter 3 for the Γ_i boundary components to have either 0 or never vanishing second derivatives, we have upper and lower bounds on the curvature of ∂D:

$$0 < \kappa_{\min} \leq \kappa \leq \kappa_{\max} < \infty \tag{3.1}$$

From (2.12) and the above mentioned fact (3.1), it is evident that if a wave front has slope $B > 0$, then $\forall t > 0$, $B_t > 0$ as well, meaning that dispersing wave fronts remain dispersing in the future.

In the case of dispersing billiards, $R = \dfrac{2\kappa}{\cos\varphi}$ is bounded away from zero: $R \geq R_{\min} := 2\kappa_{\min} > 0$ (because of our definition of a dispersing billiard).

A flat front $B = 0$ will remain flat until the first collision with ∂D, but upon that collision $B_t > 0$ and will remain dispersing. If we were to graph B_t, it would look like a sawtooth like curve: at every collision, it increases by $\dfrac{2\kappa}{\cos\varphi}$, then decreases along a hyperbola with horizontal asymptote $B = 0$. [1] has a great figure for this as well:

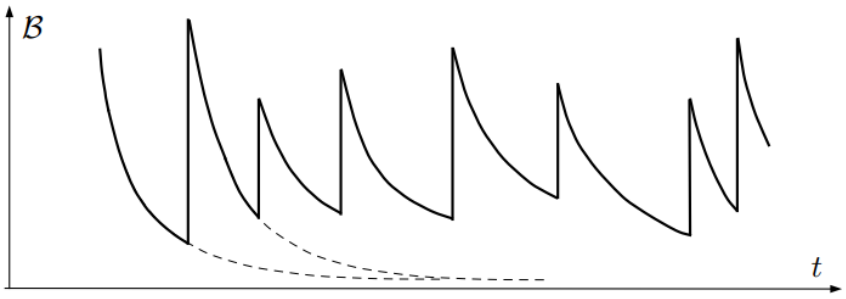

Figure 3.1: An example of how B_t might change for a dispersing billiard (symbol $\mathcal{B}$ is used instead of B_t). Figure 4.4. in [1]

Let us remember Definition 2.4.1 and the results following it. We can construct unstable cones C_x^u for $\mathcal{F}$: All wavefronts that are dispersing or flat upon arriving at x make a cone $C_x^u = \{(dr, d\varphi) \in \mathcal{T}_x\mathcal{M} : \kappa \leq \dfrac{d\varphi}{dr} \leq \infty\}$. Note how

$$\nu(x_1) = \frac{d\varphi}{dr} = B_-(x_1)\cos\varphi_1 + \kappa_1 = B_+(x_1)\cos\varphi_1 - \kappa_1$$

so it is enough to search for invariant cones for either of the quantities B_- or B_+ as well, since a lower and upper bound for these quantities also imply lower and upper bounds for ν.

From (2.12) and (3.1), C_x^u is obviously invariant, however we need strict invariance, so we have to find appropriate upper and lower bounds for these values.

For every point $x_0 \in \mathcal{M}$, we can find the upper edge ν_{up} and lower edge ν_{lo} of the cone $D_{x_0}\mathcal{F}(C_{x_0}^u) \subset T_{x_1}\mathcal{M}$ at $x_1 = \mathcal{F}(x_0)$.

If we compute the upper and lower bounds for the quantity B_-, assuming $0 < B_-(x_0) \leq \infty$ we get

$$B_{\text{lo}} = \frac{1}{2}\frac{1}{\tau_{\max} + \dfrac{1}{2\kappa_{\min}}} < \frac{1}{\tau_{\max} + \dfrac{1}{2\kappa_{\min}}} \leq \frac{1}{\tau_0 + \dfrac{\cos\varphi_1}{2\kappa_0}} \leq B_-(x_1) \leq \frac{1}{\tau_0} < \frac{1}{\tau_{\min}} + \eta = B_{\text{up}}$$

Where $\kappa_{\min}$ is the lowest-curvature boundary component, $\tau_{\min}$ and $\tau_{\max}$ are the shortest and longest distances between collisions (we know a $\tau_{\max}$ exists, because we have bounded horizon) and η is just some arbitrary positive quantity (adding this value guarantees that equality cannot hold).

From the above, the invariant cones for ν are:

$$\nu_{\text{lo}} = \kappa_{\min} < \kappa_1 + \frac{\cos\varphi_1}{\tau_0 + \dfrac{\cos\varphi_1}{2\kappa_0}}, \qquad \nu_{\text{up}} = \kappa_{\max} + \frac{1}{\tau_{\min}} > \kappa_1 + \frac{\cos\varphi_1}{\tau_0}$$

giving $\nu_{\text{lo}} < \nu \leq \nu_{\text{up}} < \kappa_{\max} + \dfrac{1}{\tau_{\min}} + \eta' < \infty$ for some arbitrary positive η', meaning our cones C_x^u are strictly invariant: $D_{x_0}\mathcal{F}(C_{x_0}^u) \subset \text{int}C_{x_1}^u \cap \{0\}$. This means the map $\mathcal{F}$ is hyperbolic.

From now on, we will work with B_- instead of ν for simplicity, and we take as granted that upper and lower bounds for B_- imply the upper and lower bounds for ν as well.

3.2 Stadia and circular billiards

In [3] L. Bunimovich points out that dispersing and defocusing are the only mechanisms of hyperbolicity that occur in billiards.

This can happen in two ways: Almost all focusing fronts must be eventually transformed into dispersing fronts by the dynamics of the system (i.e. through defocusing), or dispersion must overcome focusing so B_t grows to infinity as $t \to \infty$. In the case of dispersing billiards, the second case applied (as there were no focusing components at all). In the following problem, the first case applies.

Imagine a stadium-shaped billiard, defined by a rectangle with sidelengths d and $2R$ and 2 semicircles with radius R attached on two opposite sides of it.

In the analysis of this type of billiard, a useful trick is often applied: whenever a real trajectory collides with a flat side Γ_j on D, we can reflect the entire table across Γ_j, and let the trajectory move straight into the mirror image of D (we 'unfold' the trajectory into an infinite billiard with no flat sides).

From this trick, or through simply plugging $\kappa = 0$ into (2.11), we can see that collision with a flat wall does not change the curvature of a wavefront hitting it.

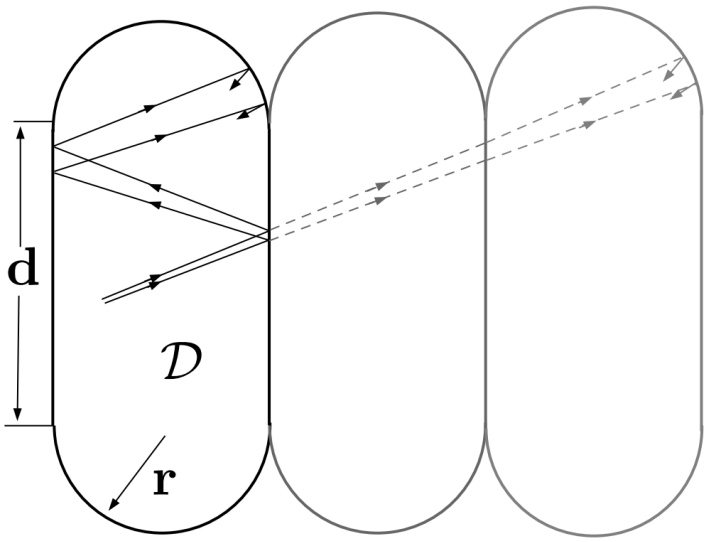

Figure 3.2: 'Unfolding' a stadium billiard (symbols $\mathcal{D}$, d and $\mathbf{r}$ are used instead of D, d and R). Figure 8.7. in [1]

Let us first investigate the $d = 0$ case to explore if there are invariant cones in a circular billiard.

In Figure 3.3, let φ be the angle of incidence of a trajectory at the boundary. Between collisions, there is a distance of $\tau = 2R\cos\varphi = 2d$.

At the next collision, it will obtain $R = -\dfrac{2}{R\cos\varphi} = -\dfrac{2}{d}$ curvature, meaning that if

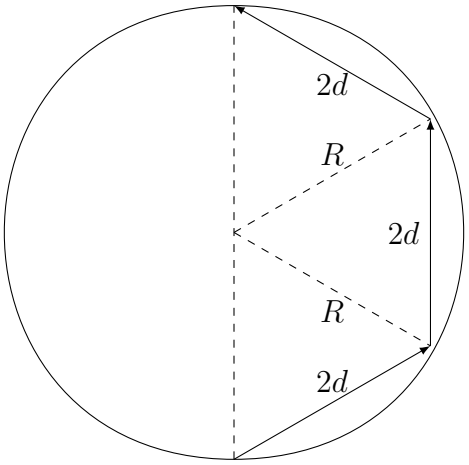

Figure 3.3: The circular billiard

its curvature before a collision was B_0^-, then

$$B_1^- = \frac{1}{2R\cos\varphi + \dfrac{1}{B_0^- - \dfrac{2}{R\cos\varphi}}} = \frac{1}{2d} - \left(\frac{1}{2d}\right)^2 \frac{1}{B_0^- - \dfrac{3}{2d}}$$

Note that

$$B_1^+ = B_1^- - \frac{2}{R\cos\varphi} = -\frac{3}{2d} - \left(\frac{1}{2d}\right)^2 \frac{1}{B_0^- - \dfrac{3}{2d}}$$

i.e. B_1^+ is 'more negative' than B_1^-, meaning that the wavefront gets focused with each collision, and it could only become dispersing through free flow. The graph of B_1^- in terms of B_0^- can be seen on Figure 3.4 below.

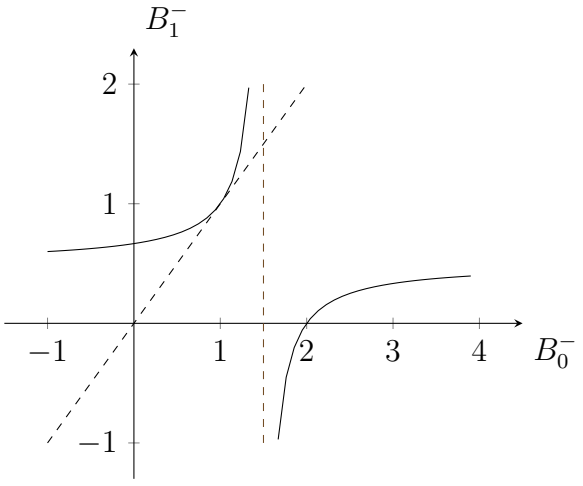

Figure 3.4: B_1^- in terms of B_0^- (without loss of generality we take $d = 1$). The vertical dashed line is at $B_0^- = \dfrac{3d}{2}$, while the other dashed line is the $y = x$ line.

$B_0^- = B_1^- = \frac{1}{d}$ describes an invariant line for the tangent map under $\mathcal{F}$. Let's analyze the cases we have:

For $B_0^- < \frac{1}{d}$, we get $B_0^- < B_1^- < \frac{1}{d}$.

For $\frac{3}{2d} < B_0^-$, we get $B_1^- < \frac{1}{2d}$, which is smaller than $\frac{1}{d}$, so it will also approach $\frac{1}{d}$ from here on.

For $\frac{1}{d} < B_0^- < \frac{3}{2d}$, we can put $B_0^- = \frac{1}{d} + \varepsilon_0$ with $0 < \varepsilon_0 < \frac{1}{2d}$ into the above formula to see that $B_1^- = B_0^- + \varepsilon_1$, where $\varepsilon_1 > \varepsilon_0$ and the difference between ε_k and ε_{k-1} also grows as the number of iterations k increases, meaning that we eventually arrive at the case $\frac{3}{2d} < B_0^- + \varepsilon_n = B_n^-$, for which the previous case applies.

This means that we can construct an invariant cone by

$$0 < B_0^- < \frac{1}{d}$$

for which

$$B_0^- < B_1^- < \frac{1}{d}$$

holds, but there is no strict invariance, since B_n^- is a monotonically increasing function in n with limit $\frac{1}{d}$, so the upper edge of the cone, $\frac{1}{d}$ does not get transformed into some value in $\operatorname{int}\left(\left[B_0^-, \frac{1}{d}\right]\right)$. Trying to extend the cone beyond $\frac{1}{d}$ will transform the values greater than $\frac{1}{d}$ out of the cone.

There is no hyperbolicity in the circular billiard. Examining the phase-space $\mathcal{M} = \left\{(r, \varphi) : r \in [0, 2R\pi), \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right\}$ (r a cyclic coordinate), we may notice that because the value of the angle of incidence φ stays constant as we iterate $\mathcal{F}$, the phase-space is foliated by lines L_φ of constant φ such that $\mathcal{F}(L_\varphi) = L_\varphi$, i.e. invariant curves, for which $\mathcal{F}$ can be interpreted as just a rotation of the circle by $\pi - 2\varphi$ (in other terms, 'the dynamics is conjugate to rotation', but this is not important here).

In the problem of $d \neq 0$, for any $d > 0$ we get that the map $\mathcal{F}$ is hyperbolic. To prove this, consider the case where $x_1 = \mathcal{F}(x_0)$ is on the other semicircle from x_0 (we know that collision with a flat component does not matter, and there is no strict invariance when viewing the same circle). Let the cone at x_0 be defined by

$$0 < B_0^- < \frac{1}{R\cos\varphi_0}$$

The distance between between the points x_0 and $\mathcal{F}(x_0) = x_1$ is $\tau = 2R\cos\varphi + \varepsilon$, where τ is bounded by $\tau_{\min} = d$ from below and $\tau_{\max} = 2R + d$ from above and ε is bounded

by $\varepsilon_{\min} = \sqrt{R^2 + d^2} - R$ from below and $\varepsilon_{\max} = d$ from above – if we allow collisions on the flat sides in between the circular arcs, there is no upper bound on τ and ε. Such „bouncing trajectory segments" are another soure of non-uniform hyperbolicity, analogous to sliding along the semicircular arc. We provide Figure 3.5 for clarity.

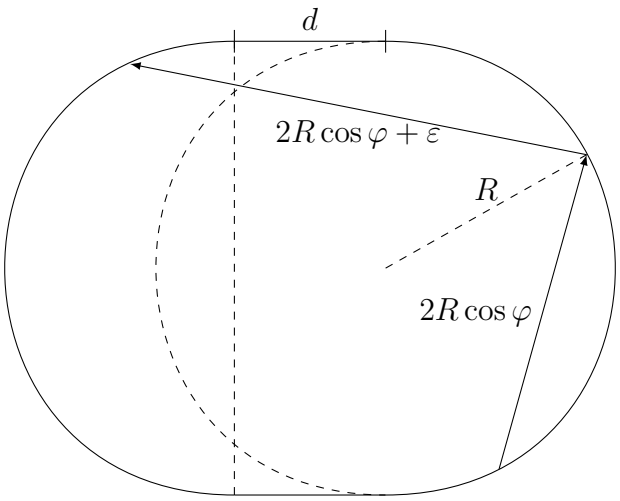

Figure 3.5: ε is the 'extra distance' of the trajectory compared to the circular billiard

From this, we get the bounds for B_1^- as:

$$0 < \frac{1}{\frac{3R}{2} + \varepsilon_{\max}} < B_1^- = \frac{1}{2R\cos\varphi_0 + \varepsilon + \frac{1}{B_0^- - \frac{2}{R\cos\varphi_0}}} < \frac{1}{R\cos\varphi_0 + \varepsilon_{\min}} < \frac{1}{R\cos\varphi_0}$$

Note that this cone's upper bound depends on the angle φ_0, and if for example $\varphi_0 > \varphi_1$, the entire cone at x_0 will not get transformed into the interior of the cone at x_1. However, in the family of billiards that the stadium billiard falls under (Bunimovich billiards), it can be proven that these cones will still result in hyperbolicity (for the proof of this, see [1] Section 8.4., Theorem 8.9.).

This means that we have found a strictly invariant cone for any x_0 and φ_0, i.e. $D_{x_0}\mathcal{F}C^u_{x_0} \subset \text{int}C^u_{\mathcal{F}(x_0)} \cup \{0\}$. In the case of the circular billiard, the upper edge of the cone is given by $\frac{1}{R\cos\varphi}$ as $\varepsilon = 0$ and φ is unchanging. In that case, the upper bound for B_1^- is going to be the same as for B_0^-, so the cone is never strictly invariant.

3.3 Annular billiards

In [4], the authors study annular tables between two circles, defined through the parameters (r, δ), where the outer circle is of radius $R = 1$, the radius of the inner circle is r, and δ is the distance between their centers. Let Ω denote the arclength of the outer circle, φ the angle of incidence of trajectories arriving on it at a given value Ω. Similarly, ω and ψ denote the arclength and angle of incidence on the smaller circles. Also denote the phase-spaces by $\mathcal{M}_{\text{out}} = \left\{(\beta, \varphi) : \beta \in [0, 2\pi), \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right\}$ and $\mathcal{M}_{\text{in}} = \left\{(\omega, \psi) : \omega \in [0, 2r\pi), \psi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$ – note that β and ω are cyclic and $0 < r$, $0 \leq \delta$ and $r + \delta < 1$ have to hold.

The construction and the parameter space of the problem is demonstrated in Figure 3.6.

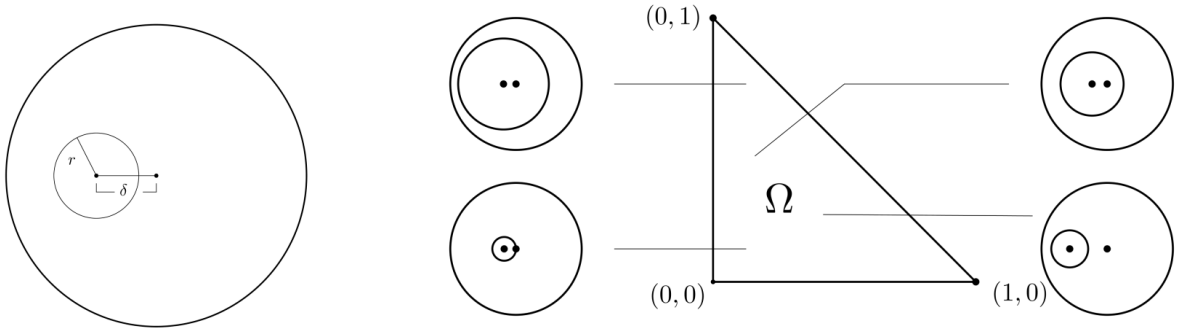

Figure 3.6: The annular table and the parameter space $\delta \times r$. Figure 1 in [4]

In this type of billiard, lots of phenomena emerge. One of the most important of these is elliptical islands for certain parameters of r and δ (elliptical islands are positive measure subsets of the phase-space which the map $\mathcal{F}$ leaves invariant, and as in the circular billiard, the dynamics is conjugate to rotation. The phase-space diagrams of these for different configurations of r and δ can be seen on Figure 3.7.

Notice how for concentric circles, the system is integrable, the entire phase-space is foliated by invariant curves C such that $\mathcal{F}(C) \subseteq C$ - in this case, these invariant curves are lines L_φ of constant φ. Also notice how for every nonconcentric billiard, the $\mathcal{M}_{\text{out}}$ phase-space has a band at the top and bottom of its phase-space that are foliated by L_φ lines that are left invariant by $\mathcal{F}$. These correspond to 'sliding' trajectories on the outer circle that avoid the inner circle entirely by having a high enough angle of incidence – these parts of the phase-space are called 'the whispering gallery'.

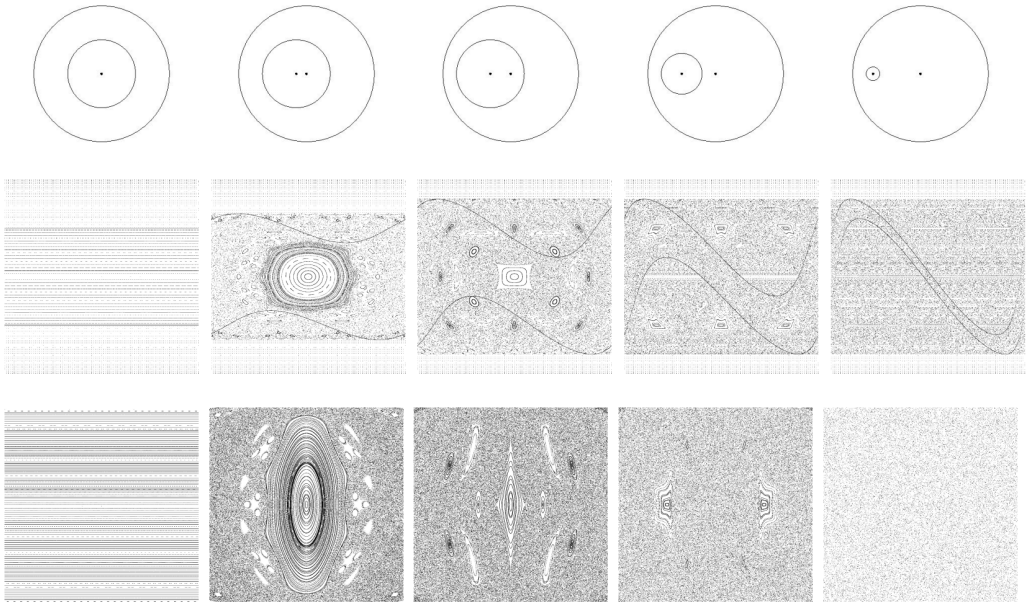

Figure 3.7: The 'typical' phase-spaces for the outside circle $\mathcal{M}_{\text{out}}$ (top) and inside circle $\mathcal{M}_{\text{in}}$ (bottom) of the problem for different values of δ, r (arclength is the horizontal axis, angle of incidence is the vertical axis). Figure 2 in [4].

For increasing values of δ and decreasing values of r, the system becomes 'more hyperbolic' – what we mean by this is that the elliptical islands appearing in both $\mathcal{M}_{\text{out}}$ and $\mathcal{M}_{\text{in}}$ shrink – but there always exist elliptical islands for any values of r, δ that satisfy our assumptions of $r + \delta < 1$.

The authors of the paper aim to show how to calibrate r and δ such that hyperbolicity occurs in a large part of the phase-space – previous results already known to them indicated that this only occurs when $\delta \approx 1$ and $r \approx 0$, i.e. small obstacle, large eccentricity cases – they find regions of hyperbolicity using the cone technique already demonstrated in previous sections. However, these hyperbolic regions they manage to construct have zero Lebesgue measure as they are fractal in nature.

In Chapter 4 of this work, our aim is also to find parameters that guarantee hyperbolicity, but instead of a zero-measure set, we obtain a full measure set of such points. Our results are not the most optimal parameters for the task, but that is not the purpose of the work. Rather, our intention with the next chapter is to demonstrate the cone technique in detail.

Chapter 4

Stadium with inner obstacle

Motivated by circular annular billiards discussed in Chapter 3 and [4], and the stadium billiard also discussed in Chapter 3, we might start to imagine a combination of the two.

Let us construct a horizontal stadium billiard with radius R semicircles, and midlength d. Then, let's imagine a circular obstacle with radius r inside the table – for sake of simplicity, let this obstacle be on the horizontal axis of symmetry. Then, „stretch" this circle horizontally into a stadium, with length d between the centers of the two semicircles, same as in the outer stadium.

This case is a special case described in [5], which investigates 'track billiards': billiards that consist of regions of annulus with circles of radii $0 < r_1 < r_2$ (circular guides) separated by rectangular regions of width $r_2 - r_1$ (straight guides).

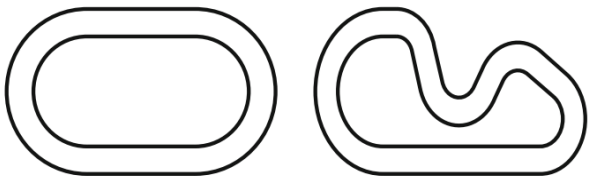

Figure 4.1: Examples of 'track billiards'. Figure 2 in [5].

What the authors of [5] conclude is that for hyperbolicity it is needed to either have long enough straight guides or wide enough and long enough circular guides – i.e. if these conditions (not discussed by us here) are not satisfied, we cannot guarantee hyperbolicity. In track billiards, another interesting phenomenon emerges: the phase-space separates into two disjoint sets which the collision map leaves invariant, corresponding to clockwise and counterclockwise directed trajectories.

This separation into two – clockwise and counterclockwise – invariant components persists whenever the centers of the semicircular arcs of the inner and outer obstacle coincide. Such a splitting occurs even in the degenerate case when the inner obstacle is just a straight segment connecting the centers of the two circular arcs. In the following, we will investigate a different, less symmetrically positioned inner obstacle, expected to result in hyperbolic and ergodic dynamics with respect to the natural invariant measure. Our discussion below only investigates hyperbolicity only, ergodicity requires further study beyond the scope of this work.

Let us construct a more generalized inner stadium: after creating the obstacle, further stretch or shrink its length by ε_1 on the right and ε_2 on the left, noting that these values' signs are such that they „point outward", meaning that if these $\varepsilon_{1,2}$ parameters are nonnegative, the distance between the inner semicircles is greater than or equal to d.

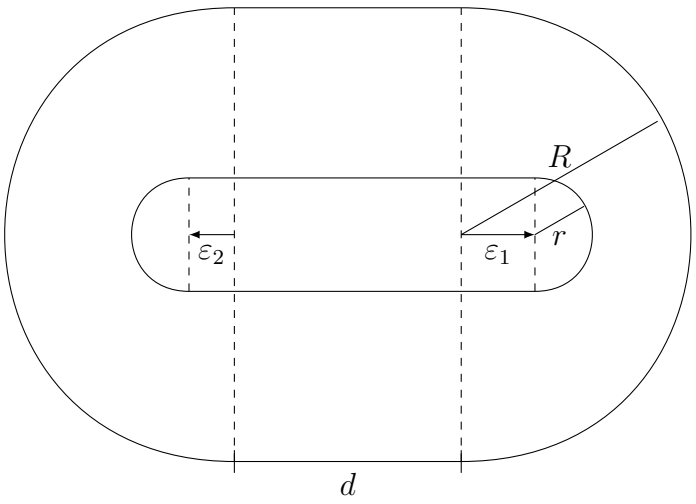

Figure 4.2: The stadium billiard with an inner obstacle, with middle section length d, outer radius R, inner radius r, and parameters $\varepsilon_1, \varepsilon_2 > 0$

It is important to note that for the sake of simplicity, we omit the investigation of the case where $d + \varepsilon_1 + \varepsilon_2 < 0$. This would lead to the semicircles overlapping (or, if we don't take the entire semicircles, corners). Thus, we already pose a restriction:

$$d + \varepsilon_1 + \varepsilon_2 \geq 0 \tag{4.1}$$

For similar reasons, we also omit the investigation of the case where $\varepsilon_1 + r$ or $\varepsilon_2 + r$ are greater than or equal to R, in which case either the inner semicircle would be tangent to the outer semicircle, or the inner boundary would stick out through the outer boundary

(this way, 'inner' and 'outer' would no longer be meaningful specifications of the curves either).

4.1 Hyperbolicity between collisions

Collision with the flat components of the billiard table can be omitted from the discussion using a 'mirroring' method similarly to stadia in Chapter 3. This requires the flat sides of the billiard table to be 'glued' together (inner flat side with other inner flat side, outer flat side with other outer flat side), in the sense that a particle arriving at one of the flat sides will continue without reflection from the flat side it is 'glued' to, keeping its x coordinate, but jumping in its y coordinate by $\pm 2r$ (depending on which direction the trajectory started in). Then, because we placed our obstacle on the axis of horizontal symmetry, this is essentially the same as if we flipped the billiard along the flat side as soon as the trajectory arrived at it.

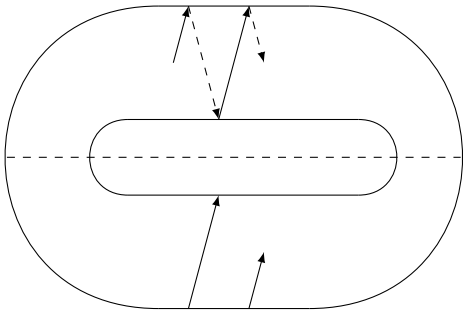

Figure 4.3: The trajectory of a particle arriving on the 'glued together' flat sides in the problem. The dashed lines together with the solid lines in the upper region represent the real trajectory of the billiard, while the collection of solid lines represents the imaginary trajectory of the billiard, which never gets reflected on flat components. The solid lines in the lower region are mirror images of the dashed lines in the upper region.

This convention for gluing is enough for our purposes, as we are only interested whether the system is hyperbolic, in which case what only matters is the distance travelled between collisions with curved components. If we were interested whether the system is ergodic or not, or whether the phase-space separates into clockwise and counterclockwise trajectories like in track billiards, we would have to carefully construct a rather hard-to-imagine 2D-surface without flat components, on which the flow preserves clockwise and counterclockwise motion.

It is enough to prove strict invariance of the cone when the two consecutive collisions are on two different curved boundary components (with no bounces on the flat components in between). For such transitions, there is a uniform upper bound on the free path, which we denote below by $\tau_{\max}$. In the case of the stadium billiard, $\tau_{\max} = d + 2R$, in our problem, $\tau_{\max} < d + 2R$ between any curved boundary component if the inner obstacle has nonzero thickness, i.e. $r > 0$.

Note that trajectories that consist only of collisions with the flat parts of the boundary also exist, they are even dense inside the allowed parts of the rectangle, but it can be easily verified that they have measure 0 on the phase space of the billiard map, just like trajectories colliding with a corner at the end of the flat sides.

Our next goal is to find the conditions that guarantee hyperbolicity when going from one curved part of the boundary to the other. Consider the labeling in Figure 4.4. The unique cases for trajectories between two curved parts of the boundary are listed below:

- $\Gamma_1 \to \Gamma_1$
- $\Gamma_1 \to \gamma_1$
- $\Gamma_1 \to \gamma_2$
- $\Gamma_1 \to \Gamma_2$
- $\gamma_1 \to \Gamma_1$
- $\gamma_1 \to \Gamma_2$

Note that in each case above, the indices 1 and 2 can be swapped, and listing every case this way will result in all possible curve-to-curve trajectories.

4.2 Hyperbolicity at the outer boundary

We want to prove hyperbolicity by finding invariant cones for B^- through the following:

$$C_x^u = \{(d\xi, d\omega) \in \mathcal{T}_x\Omega \mid \exists c_1, c_2 \geq 0 : \quad c_1 < B^-(x) = \frac{d\xi}{d\omega} < c_2\}$$

$$D_x\mathcal{F}(C_x^u) \subset \text{int}C_{\mathcal{F}(x)}^u \cup \{0\}$$

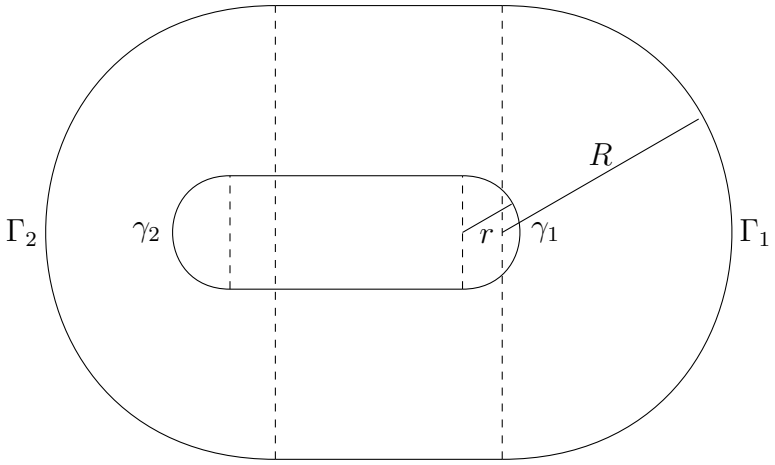

Figure 4.4: Γ_1 and Γ_2 denote the outer semicircles, γ_1 and γ_2 denote the inner ones

Assume that a wavefront with curvature B_1^- starts just before colliding with Γ_1 and gets to the curve γ_1, γ_2 or Γ_2 in the next collision (the same reasoning also works for switching Γ_1 and Γ_2).

Let the cone at $x_1 \in \Gamma_1$ be defined by

$$0 < B_1^-(x_1) < \frac{1}{R},$$

and let $x_2 = \mathcal{F}x_1$. Then – as we will argue below – the image of this cone at $x_2 \in \gamma_1, \gamma_2$ or Γ_2 respectively can be given by

$$\frac{1}{\tau_{\max}} < B_2^-(x_2) < \frac{1}{\tau_{\min} - R} \tag{4.2}$$

where $\tau_{\min}$ is different for the different curves $\gamma_1, \gamma_2, \Gamma_2$.

After the collision with Γ_1 at x_1, the curvature will have become

$$B_1^+ = B_1^- - \frac{2}{R\cos\varphi} = B_1^- - \beta$$

where

$$\frac{2}{R} \leq \beta = \frac{2}{R\cos(\varphi)} < \infty \tag{4.3}$$

(„$< \infty$" here means that grazing collisions on the outer boundary are not possible, unlike with the inner boundary, where equality will also be possible). Then, just before the collision with the next curve (at $x_2 \in \gamma_1, \gamma_2$ or Γ_2), the curvature that started out as B_1^- at $x_1 \in \Gamma_1$ will become

$$B_2^- = \frac{1}{\tau + \dfrac{1}{B_1^- - \beta}} \tag{4.4}$$

where

$$0 < \tau_{\min} \leq \tau < \tau_{\max} \tag{4.5}$$

Let us construct result (4.4) step-by-step while attempting to find an invariant cone as in (4.2) above. To start, let us assume only that $0 < B_1^-$. Then, after having bounced off of the outer boundary, the curvature will have changed:

$$-\infty < B_1^+ = B_1^- - \beta \leq B_1^- - \frac{2}{R}$$

For the construction of the invariant cone, we want to avoid the situation where $B_1^- - \frac{2}{R} \geq 0$ (as that would lead to $1/B_1^+$ falling into one of 2 disjoint intervals, one with positive lower bound, the other with negative upper bound, both of these intervals unbounded on their other sides), so we already put a further restriction on B_1^-:

$$0 < B_1^- < \frac{2}{R}$$

Now, take the reciprocal of the expression for B_1^+:

$$\frac{1}{B_1^- - \frac{2}{R}} \leq \frac{1}{B_1^+} = \frac{1}{B_1^- - \beta} < 0$$

Adding τ to $1/B_1^+$:

$$\tau_{\min} + \frac{1}{B_1^- - \frac{2}{R}} \leq \frac{1}{B_2^-} = \tau + \frac{1}{B_1^- - \beta} < \tau_{\mathbf{max}}$$

Again, for constructing the invariant cone, we want to omit the case when $\tau_{\min} + \frac{1}{B_1^- - \frac{2}{R}} \leq 0$. The previous restriction for B_1^- is not enough to make such a lower bound, but for example

$$0 < B_1^- < \frac{1}{R}$$

works, and will result in the restriction

$$\tau_{\min} > R \tag{4.6}$$

The above restrictions will result in the following lower and upper bounds for B_2^-:

$$\frac{1}{\tau_{\max}} < B_2^- = \frac{1}{\tau + \frac{1}{B_1^- - \beta}} < \frac{1}{\tau_{\min} + \frac{1}{B_1^- - \frac{2}{R}}} < \frac{1}{\tau_{\min} - R} \tag{4.7}$$

Thus, we have found positive lower and upper bounds for B_1^- and $\tau_{\min}$ such that there exist more restrictive positive lower and upper bounds for B_2^- as well – thus we have found an invariant cone. Note that we have strict invariance as well, because of the finiteness of these bounds. Also note that the upper edge of the cone could have been $\frac{1}{R\cos\varphi}$ as well, in which case we still have strict invariance, and it would be in accordance with the circular billiard's invariant, but not strictly invariant cone – that would pose the restriction $\tau_{\min} > R\cos\varphi$, which is always satisfied by $\tau_{\min} > R$.

4.3 Hyperbolicity at the inner boundary

Assume a wavefront with curvature B_1^- starts just before colliding with γ_1 and gets to the curve Γ_1 or Γ_2 (the same reasoning works when starting from γ_2 as well). For this case, we will let the cones be defined by the following:

$$0 < B_1^-(x_1) \leq \infty$$
$$\frac{1}{2\tau_{\max}} < \frac{1}{\tau_{\max}} \leq B_2^-(x_2) \leq \frac{1}{\tau_{\min}} < \frac{2}{\tau_{\min}}$$

Just before the collision with the next curve at $x_2 \in \Gamma_1$ or Γ_2, the cruvature that started out as B_1^- at $x_1 \in \gamma_1$ will become

$$B_2^- = \frac{1}{\tau + \frac{1}{B_1^- + \beta}} \tag{4.8}$$

with

$$\frac{2}{r} \leq \beta = \frac{2}{r\cos(\varphi)} \leq \infty \tag{4.9}$$

and (4.5) still holds (perhaps with different $\tau_{\min}$ and $\tau_{\max}$ values). Again, let us construct the result step by step with the assumption $B_1^- > 0$:

$$\frac{2}{r} < B_1^- + \beta \leq \infty$$

Taking the reciprocal:

$$\frac{r}{2} \geq \frac{1}{B_1^- + \beta} \geq 0$$

Adding τ:

$$\tau_{\min} \leq \tau + \frac{1}{B_1^- + \beta} < \tau_{\max} + \frac{r}{2}$$

Taking the reciprocal, we finally get

$$\frac{1}{\tau_{\max}} < \frac{1}{\tau + \dfrac{1}{B_1^- + \beta}} \leq \frac{1}{\tau_{\min}} < \frac{1}{\tau_{\min}} + \eta \tag{4.10}$$

for arbitrary positive η. However, this cone needs to be inside the cone discussed in the previous section: $0 < \tau_{\max}$ trivially holds, but $\dfrac{1}{\tau_{\min}} - \eta < \dfrac{1}{R} - \eta'$ also has to hold for some small but positive η, η' values – luckily, the previous $\tau_{\min} > R$ restriction guarantees that there exist such η, η'. Thus, we have found an invariant cone without posing any new restriction for B_1^- or $\tau_{\min}$.

4.4 Parameters guaranteeing hyperbolicity

Reminding ourselves that (4.7) and (4.10) work for trajectories starting at not only Γ_1 and γ_1, but also at Γ_2 and γ_2, now we aim to find the collection of parameters $(d, r, \varepsilon_1, \varepsilon_2)$ that guarantee hyperbolicity between collisions with the curved boundaries. In the case resulting in (4.10) we do not need to pose any restrictions for the parameters. For the case resulting in (4.7) with the restriction (4.6), we will go through the unique cases that start with Γ_1, mentioned in Section 4.1.

As we've seen with the circular billiard, in the $\Gamma_1 \to \Gamma_1$ case there is an invariant cone present, but it's not a strictly invariant cone, so there is no hyperbolicity (also, $\tau_{\min} = 0$ independently of the parameters, thus $\tau_{\min} \not> R$). However, it's evident that after 'enough' collisions, the trajectory will leave the arc, then after some τ distance, collide with either γ_1, γ_2 or Γ_2, so all that's left is to examine these cases.

We want to find $\tau_{\min}$ in all cases and ensure that it's greater than R. For this, consider 2 general nonintersecting circular arcs. Then, the smallest distance between the two comes from one of the following sets of line segments that connect **a**) their endpoints, **b**) an endpoint of one and an interior point of the other, such that this line segment's extension goes through the centre of the second arc, **c**) interior points of the two arcs, such that this line segment's extension goes through the centres of the arcs.

Let $R, r > 0$ be fixed, and examine the cases $\Gamma_1 \to \gamma_1$ depending on the sign of ε_1, with the help of the figures below.

For the case of $\varepsilon_1 > 0$, $\tau_{\min} = R - r - \varepsilon_1 > R$ has no positive solution for ε_1, meaning that in this case we can't guarantee hyperbolicity.

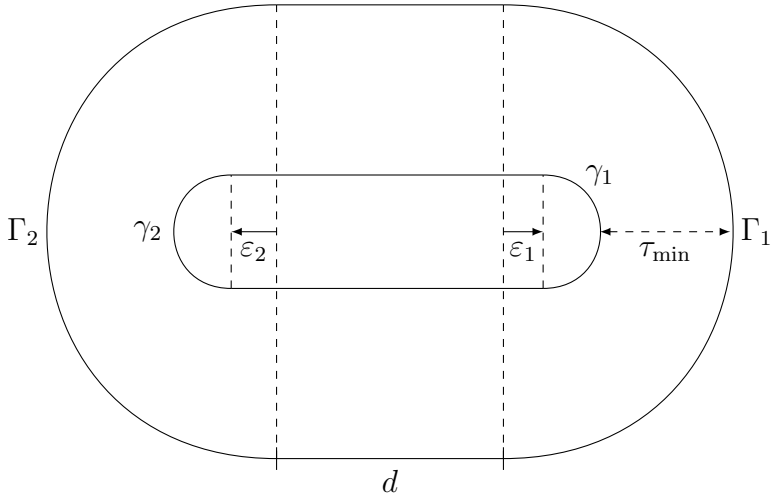

Figure 4.5: Case c) is valid for $\varepsilon_1 > 0$

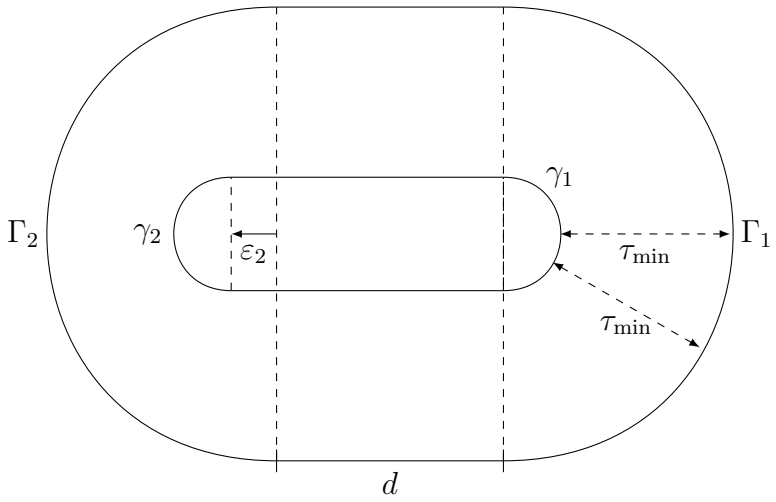

Figure 4.6: Case c) is valid for $\varepsilon_1 = 0$

For the case of $\varepsilon_1 = 0$, Γ_1 and γ_1 are concentric, so $\tau_{\min} = R - r > R$ has no solution either.

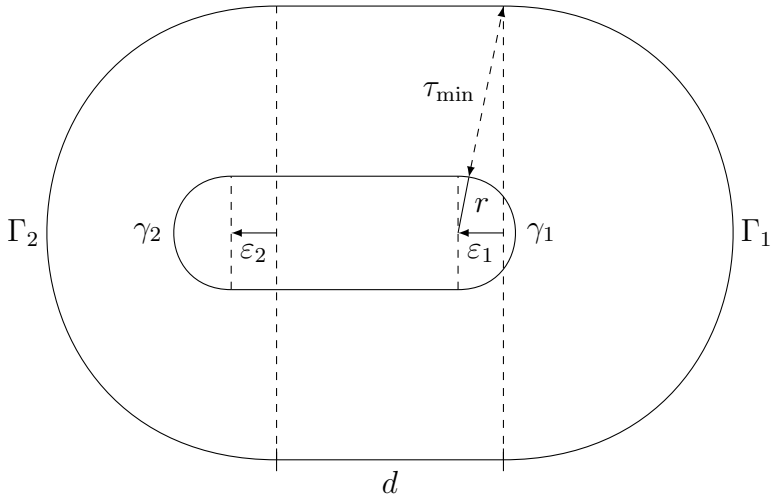

Figure 4.7: Case *b*) is valid for $\varepsilon_1 < 0$

For the case of $\epsilon_1 < 0$, $\tau_{\min} = \sqrt{R^2 + \varepsilon_1^2} - r > R$ has the solution $|\varepsilon_1| > \sqrt{r^2 + 2Rr}$ and combining this with $\varepsilon_1 < 0$, then noting that this has to hold for the $\Gamma_2 \to \gamma_2$ case as well, we get:

$$\varepsilon_{1,2} < -\sqrt{2Rr + r^2} \tag{4.11}$$

Note that for the $\Gamma_1 \to \Gamma_2$ case we get $\tau_{\min} = d$ trivially, for which $d > R$ is the restriction we get, but given that we didn't aim to find the most optimal cone, just 'some cone', we neglect this restriction and refer to the case of the stadium, meaning that $d > 0$ is enough of a restriction, which we already have from (4.1) and (4.11)

Now, let's examine the $\Gamma_1 \to \gamma_2$ case, noting that we already pose restriction (4.11), implemented in the figure below as well.

Because of (4.1) and (4.11), $\varepsilon_2 < -d$ is not possible, giving us $\tau_{\min} = \sqrt{(d + \varepsilon_2)^2 + (R - r)^2} > R$, which has the solution

$$|d + \varepsilon_2| = d + \varepsilon_2 > \sqrt{2Rr - r^2}$$

Then, taking (4.1), we get

$$d + \varepsilon_2 > -\varepsilon_1 > \sqrt{2Rr + r^2} > \sqrt{2Rr - r^2}$$

so we get no new restriction for $\varepsilon_{1,2}$.

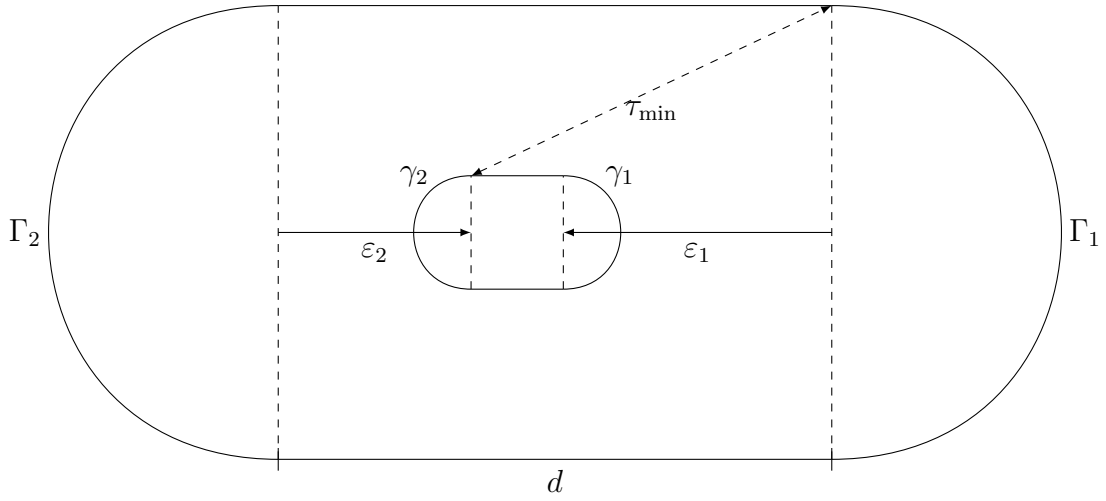

Figure 4.8: Case a) is valid for $-d < \varepsilon_2 < 0$

Thus now, we have found the space of values $(d, \varepsilon_1, \varepsilon_2)$ for any $R > r > 0$ that guarantee hyperbolicity:

$$\begin{cases} \varepsilon_1 < -\sqrt{2Rr + r^2} \\ \varepsilon_2 < -\sqrt{2Rr + r^2} \\ d > -\varepsilon_1 - \varepsilon_2 \end{cases} \tag{4.12}$$

Note however, that these are not the most optimal values of these quantities that guarantee hyperbolicity overall, not even the most optimal values that guarantee hyperbolicity between any collisions from one curved boundary to the other. The aim of this derivation was to demonstrate the calculation for finding cones for such an unusual billiard table.

Note, that if we had viewed a series of collisions instead of just one, we could have got much more optimal values. However, for example, viewing the period 2 trajectory $\gamma_1 \to \Gamma_1 \to \gamma_1$, we get

$$\frac{1}{\tau_{\max}} < B_2^- < \frac{1}{\tau_{\min} + \dfrac{1}{\dfrac{1}{\tau_{\min}} - \dfrac{2}{R}}}$$

And if we again only care for 'some cone', not the most optimal one, the same $\tau_{\min} > R$ restriction is satisfactory.

It is important to note that there are specific conditions discussed in general works on hyperbolicity [7] and ergodicity [6] for some classes of billiards which cover this example, but rather than checking the conditions laid out in these works, we aimed to construct the invariant cones directly.

Chapter 5

Conclusion

In this work, we introduced the basics of billiard dynamics using a very general book [1] as our main source of information. We used it to study dispersing, stadium-shaped and circular billiards, then we summarized the relevant results of [4]. After these, we worked on the unique problem of a stadium billiard with an inner obstacle, and found a family of parameters that guarantee hyperbolicity between 2 consecutive collisions on different arcs of the boundary.

In the future, further work could be put into the problem to find more optimal parameters for hyperbolicity, find parameters that guarantee ergodicity – perhaps, in another direction generalize the results known for the $\varepsilon_1 = \varepsilon_2 = 0, r > 0$ case, i.e. the separation of the phase-space into 2 ergodic components, one corresponding to clockwise directed trajectories, while the other corresponding to counterclockwise directed trajectories.

The author is also interested in studying quantum chaos in the future (these are quantum systems whose classical equivalent is chaotic), for which writing this bachelor thesis was rather useful, as he had to familiarize himself with the relevant concepts and literature required to examine classically chaotic systems.

Bibliography

[1] Nikolai Chernov, Roberto Markarian, *Chaotic Billiards*, Mathematical Surveys and Monographs of the AMS, vol. 127, 2006, ISBN: 0821840967

[2] M.P.Wojtkowski *Invariant families of cones and Lyapunov exponents*, Ergodic Theory and Dynamical Systems, 5, 145–161, 1985

[3] Leonid A Bunimovich *Mechanisms of chaos in billiards: dispersing, defocusing and nothing else*, Nonlinearity 31 R78, 2018

[4] R B Batista, M J Dias Carneiro, S Oliffson Kamphorst *Hyperbolicity and Abundance of Elliptical Islands in Annular Billiards*, Ergodic Theory and Dynamical Systems pp. 1-33., 33 p., 2023

[5] Bunimovich L and Del Magno G *Track billiards*, Communications in Mathemathical Physics 288 699–713, 2009

[6] Del Magno G, R Markarian*On the Bernoulli Property of Planar Hyperbolic Billiards*, Communications in Mathematical Physics 350, 917–955, 2017

[7] Donnay, V.J.*Using integrability to produce chaos: billiards with positive entropy*, Communications in Mathematical Physics 141, 225-257, 1991