Chaotic Billiards

Nikolai Chernov

Roberto Markarian

Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294, USA
E-mail address: chernov@math.uab.edu

Instituto de Matemática y Estadística "Prof. Ing. Rafael Laguardia" Facultad de Ingeniería, Universidad de la República, C.C. 30, Montevideo, Uruguay
E-mail address: roma@fing.edu.uy

To Yakov Sinai on the occasion of his 70th birthday

Contents

Preface

Billiards are mathematical models for many physical phenomena where one or more particles move in a container and collide with its walls and/or with each other. The dynamical properties of such models are determined by the shape of the walls of the container, and they 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.

Mathematical theory of chaotic billiards was born in 1970 when Ya. Sinai published his seminal paper [**Sin70**], and now it is only 35 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.

It is no surprise that many young mathematicians and scientists attempt to learn chaotic billiards in order to investigate some of these and related physical models. But such studies are often prohibitively difficult for many novices and outsiders, not only because the subject itself is intrinsically quite complex, but to a large extent because of the lack of comprehensive introductory texts.

True, there are excellent books covering general mathematical billiards [**Ta95, KT91, KS86, GZ90, CFS82**], but these barely touch upon chaotic models. There are surveys devoted to chaotic billiards as well (see [**DS00, HB00, CM03**]) but those are expository; they only sketch selective arguments and rarely get down to 'nuts and bolts'. For readers who want to look 'under the hood' and become professional (and we speak of graduate students and young researchers here), there is not much choice left: either learn from their advisors or other experts by way of personal communication or read the original publications (most of them very long and technical articles translated from Russian). Then students quickly discover that some essential facts and techniques can be found only in the middle of long dense papers. Worse yet, some of these facts have never even been published – they exist as folklore.

This book attempts to present the fundamentals of the mathematical theory of chaotic billiards in a systematic way. We cover all the basic facts, provide full proofs, intuitive explanations and plenty of illustrations. Our book can be used by students and self-learners. It starts with the most elementary examples and formal definitions and then takes the reader step by step into the depth of Sinai's theory of hyperbolicity and ergodicity of chaotic billiards, as well as more recent achievements related to their statistical properties (decay of correlations and limit theorems).

The reader should be warned that our book is designed for active learning. It contains plenty of exercises of various kinds: some constitute small steps in the proofs of major theorems, others present interesting examples and counterexamples, yet others are given for the reader's practice (some exercises are actually quite challenging). The reader is strongly encouraged to do exercises when reading the book, as this is the best way to grasp the main concepts and eventually master the techniques of billiard theory.

The book is restricted to two-dimensional chaotic billiards, primarily dispersing tables by Sinai and circular-arc tables by Bunimovich (with some other planar chaotic billiards reviewed in the last chapter). We have several compelling reasons for such a confinement. First, Sinai's and Bunimovich's billiards are the oldest and best explored (for instance, statistical properties are established only for them and for no other billiard model). The current knowledge of other chaotic billiards is much less complete; the work on some of them (most notably, hard ball gases) is currently under way and should perhaps be the subject of future textbooks. Second, the two classes presented here constitute the core of the entire theory of chaotic billiards. All its apparatus is built upon the original works by Sinai and Bunimovich, but their fundamental works are hardly accessible to today's students or researchers, as there have been no attempts to update or republish their results since the middle 1970s (after Gallavotti's book [**Ga74**]). Our book makes such an attempt. We do not cover polygonal billiards, even though some of them are mildly chaotic (ergodic). For surveys of polygonal billiards see [**Gut86, Gut96**].

We assume that the reader is familiar with standard graduate courses in mathematics: linear algebra, measure theory, topology, Riemannian geometry, complex analysis, probability theory. We also assume knowledge of ergodic theory. Although the latter is not a standard graduate course, it is absolutely necessary for reading this book. We do not attempt to cover it here, though, as there are many excellent texts around [**Wa82, Man83, KH95, Pet83, CFS82, DS00, BrS02, Dev89, Sin76**] (see also our previous book, [**CM03**]). For the reader's convenience, we provide basic definitions and facts from ergodic theory, probability theory, and measure theory in the appendices.

Acknowledgements. The authors are grateful to many colleagues who have read the manuscript and made numerous useful remarks, in particular P. Balint, D. Dolgopyat, C. Liverani, G. Del Magno, and H.-K. Zhang. It is a pleasure to acknowledge the warm hospitality of IMPA (Rio de Janeiro), where the final version of the book was prepared. We also thank the anonymous referees for helpful comments. Last but not least, the book was written at the suggestion of Sergei Gelfand and thanks to his constant encouragement. The first author was partially supported by NSF grant DMS-0354775 (USA). The second author was partially supported by a Proyecto PDT-Conicyt (Uruguay).

Symbols and notation

$\mathcal{D}$	billiard table	Section	2.1
Γ	boundary of the billiard table		2.1
Γ_+	union of dispersing components of the boundary Γ		2.1
Γ_-	union of focusing components of the boundary Γ		2.1
Γ_0	union of neutral (flat) components of the boundary Γ		2.1
$\tilde{\Gamma}$	regular part of the boundary of billiard table		2.1
Γ_*	Corner points on billiard table		2.1
ℓ	degree of smoothness of the boundary $\Gamma = \partial\mathcal{D}$		2.1
n	normal vector to the boundary of billiard table		2.3
$\mathbf{T}$	tangent vector to the boundary of billiard table		2.6
$\mathcal{K}$	(signed) curvature of the boundary of billiard table		2.1
Φ^t	billiard flow		2.5
Ω	the phase space of the billiard flow		2.5
$\tilde{\Omega}$	part of phase space where dynamics is defined at all times		2.5
π_q, π_v	projections of Ω to the position and velocity subspaces		2.5
ω	angular coordinate in phase space Ω		2.6
η, ξ	Jacobi coordinates in phase space Ω		3.6
μ_Ω	invariant measure for the flow Φ^t		2.6
$\mathcal{F}$	collision map or billiard map		2.9
$\mathcal{M}$	collision space (phase space of the billiard map)		2.9
$\tilde{\mathcal{M}}$	part of $\mathcal{M}$ where all iterations of $\mathcal{F}$ are defined		2.9
$\hat{\mathcal{M}}$	part of $\mathcal{M}$ where all iterations of $\mathcal{F}$ are smooth		2.11
r, φ	coordinates in the collision space $\mathcal{M}$		2.10
μ	invariant measure for the collision map $\mathcal{F}$		2.12
$\mathcal{S}_0$	boundary of the collision space $\mathcal{M}$		2.10
$\mathcal{S}_{\pm 1}$	singularity set for the map $\mathcal{F}^{\pm 1}$		2.10
$\mathcal{S}_{\pm n}$	singularity set for the map $\mathcal{F}^{\pm n}$		2.11
$\mathcal{S}_{\pm\infty}$	same as $\cup_{n\geq 1}\mathcal{S}_{\pm n}$		4.11
$\mathcal{Q}_n(x)$	connected component of $\mathcal{M} \setminus \mathcal{S}_n$ containing x		4.11
$\mathcal{V}$	$(= d\varphi/dr)$ slope of smooth curves in $\mathcal{M}$		3.10
τ	return time (intercollision time)		2.9
$\bar{\tau}$	mean return time (mean free path)		2.12
$\lambda_x^{(i)}$	Lyapunov exponent at the point x		3.1
E_x^s, E_x^u	stable and unstable tangent subspaces at the point x		3.1
$\mathcal{C}_x^s, \mathcal{C}_x^u$	stable and unstable cones at the point x		3.13
Λ	(minimal) factor of expansion of unstable vectors		4.4
$\mathcal{B}$	the curvature of wave fronts		3.7

$\mathcal{R}$	collision parameter	3.6
$\mathbb{H}_k$	homogeneity strips	5.3
$\mathbb{S}_k$	lines separating homogeneity strips	5.3
k_0	minimal nonzero index of homogeneity strips	5.3
$\mathcal{M}_{\mathbb{H}}$	new collision space (union of homogeneity strips)	5.4
$\mathbf{h}$	holonomy map	5.7
$\mathcal{I}$	involution map	2.14
$\mathbf{m}$	Lebesgue measure on lines and curves	5.9
$\lvert W\rvert$	length of the curve W	4.5
$\lvert W\rvert_p$	length of the curve W in the p-metric	4.5
$\mathcal{J}_W\mathcal{F}^n(x)$	Jacobian of the restriction of $\mathcal{F}^n$ to the curve W at the point $x \in W$	5.2
$r_W(x)$	distance from $x \in W$ to the nearest endpoint of the curve W	4.12
$r_n(x)$	distance from $\mathcal{F}^n(x)$ to the nearest endpoint of the component of $\mathcal{F}^n(W)$ that contains $\mathcal{F}^n(x)$	5.9
$p_W(x)$	distance from $x \in W$ to the nearest endpoint of W in the p-metric	4.13
$\rho_W(x)$	u-SRB density on unstable manifold W	5.2
$\asymp$	'same order of magnitude'	4.3
L	ceiling function for suspension flows	2.9

CHAPTER 1

Simple examples

We start with a few simple examples of mathematical billiards, which will help us introduce basic features of billiard dynamics. This chapter is for the complete beginner. The reader familiar with some billiards may safely skip it – all the formal definitions will be given in Chapter 2.

1.1. Billiard in a circle

Let $\mathcal{D}$ denote the unit disk $x^2+y^2 \le 1$. Let a point-like (dimensionless) particle move inside $\mathcal{D}$ with constant speed and bounce off its boundary $\partial\mathcal{D}$ according to the classical rule *the angle of incidence is equal to the angle of reflection*; see below.

Denote by $q_t = (x_t, y_t)$ the coordinates of the moving particle at time t and by $v_t = (u_t, w_t)$ its velocity vector. Then its position and velocity at time $t+s$ can be computed by

$$\begin{aligned} x_{t+s} &= x_t + u_t s \qquad & u_{t+s} &= u_t \\ y_{t+s} &= y_t + w_t s \qquad & w_{t+s} &= w_t \end{aligned} \tag{1.1}$$

as long as the particle stays inside $\mathcal{D}$ (makes no contact with $\partial\mathcal{D}$).

When the particle collides with the boundary $\partial\mathcal{D} = \{x^2+y^2=1\}$, its velocity vector v gets reflected across the tangent line to $\partial\mathcal{D}$ at the point of collision; see Fig. 1.1.

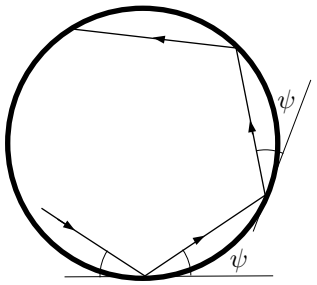

FIGURE 1.1. Billiard motion in a circle.

EXERCISE 1.1. Show that the new (postcollisional) velocity vector is related to the old (precollisional) velocity by the rule

$$v^{\text{new}} = v^{\text{old}} - 2\,\langle v^{\text{old}}, n\rangle\, n, \tag{1.2}$$

where $n = (x, y)$ is the unit normal vector to the circle $x^2+y^2=1$ and $\langle v, n\rangle = ux + wy$ denotes the scalar product.

After the reflection, the particle resumes its free motion (1.1) inside the disk $\mathcal{D}$, until the next collision with the boundary $\partial\mathcal{D}$. Then it bounces off again, and so on. The motion can be continued indefinitely, both in the future and the past.

For example, if the particle runs along a diameter of the disk, its velocity vector will get reversed at every collision, and the particle will keep running back and forth along the same diameter forever. Other examples of periodic motion are shown in Fig. 1.2, where the particle traverses the sides of some regular polygons.

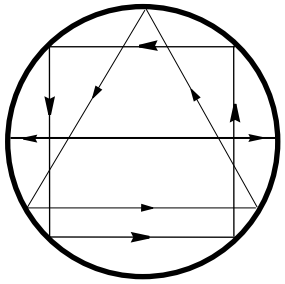

FIGURE 1.2. Periodic motion in a circle.

In the studies of dynamical systems, the primary goal is to describe the evolution of the system over long time periods and its asymptotic behavior in the limit $t \to \infty$. We will focus on such a description.

Let us parameterize the unit circle $x^2 + y^2 = 1$ by the polar (counterclockwise) angle $\theta \in [0, 2\pi]$ (since θ is a cyclic coordinate, its values 0 and 2π are identified). Also, denote by $\psi \in [0, \pi]$ the angle of reflection as shown in Fig. 1.1.

REMARK 1.2. We note that θ is actually an arc length parameter on the circle $\partial\mathcal{D}$; when studying more general billiard tables $\mathcal{D}$, we will always parameterize the boundary $\partial\mathcal{D}$ by its arc length. Instead of ψ, a reflection can also be described by the angle $\varphi = \pi/2 - \psi \in [-\pi/2, \pi/2]$ that the postcollisional velocity vector makes with the inward normal to $\partial\mathcal{D}$. In fact, all principal formulas in this book will be given in terms of φ rather than ψ, but for the moment we proceed with ψ.

For every $n \in \mathbb{Z}$, let θ_n denote the nth collision point and ψ_n the corresponding angle of reflection.

EXERCISE 1.3. Show that

$$\begin{aligned} \theta_{n+1} &= \theta_n + 2\psi_n \qquad (\mathrm{mod}\ \ 2\pi) \\ \psi_{n+1} &= \psi_n \end{aligned} \tag{1.3}$$

for all $n \in \mathbb{Z}$.

We make two important observations now:

- All the distances between reflection points are equal.
- The angle of reflection remains unchanged.

COROLLARY 1.4. *Let (θ_0, ψ_0) denote the parameters of the initial collision. Then*

$$\begin{aligned} \theta_n &= \theta_0 + 2n\psi_0 \qquad (\mathrm{mod}\ \ 2\pi) \\ \psi_n &= \psi_0. \end{aligned}$$

Every collision is characterized by two numbers: θ (the point) and ψ (the angle). All the collisions make the *collision space* with coordinates θ and ψ on it. It is a cylinder because θ is a cyclic coordinate; see Fig. 1.3. We denote the collision space by $\mathcal{M}$. The motion of the particle, from collision to collision, corresponds to a map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$, which we call the *collision map*. For a circular billiard it is given by equations (1.3).

Observe that $\mathcal{F}$ leaves every horizontal level $\mathcal{C}_\psi = \{\psi = \text{const}\}$ of the cylinder $\mathcal{M}$ invariant. Furthermore, the restriction of $\mathcal{F}$ to $\mathcal{C}_\psi$ is a rotation of the circle $\mathcal{C}_\psi$ through the angle 2ψ. The angle of rotation continuously changes from circle to circle, growing from 0 at the bottom $\{\psi = 0\}$ to 2π at the top $\{\psi = \pi\}$ (thus the top and bottom circles are actually kept fixed by $\mathcal{F}$). The cylinder $\mathcal{M}$ is "twisted upward" ("unscrewed") by the map $\mathcal{F}$; see Fig. 1.3.

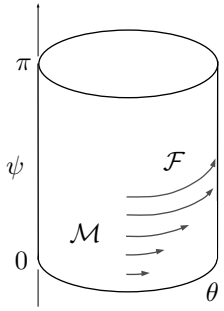

FIGURE 1.3. Action of the collision map $\mathcal{F}$ on $\mathcal{M}$.

Rigid rotation of a circle is a basic example in ergodic theory; cf. Appendix C. It preserves the Lebesgue measure on the circle. Rotations through rational angles are periodic, while those through irrational angles are ergodic.

EXERCISE 1.5. Show that if $\psi < \pi$ is a rational multiple of π, i.e. $\psi/\pi = m/n$ (irreducible fraction), then the rotation of the circle $\mathcal{C}_\psi$ is periodic with (minimal) period n, that is every point on that circle is periodic with period n, i.e. $\mathcal{F}^n(\theta, \psi) = (\theta, \psi)$ for every $0 \le \theta \le 2\pi$.

If ψ/π is irrational, then the rotation of $\mathcal{C}_\psi$ is ergodic with respect to the Lebesgue measure. Furthermore, it is *uniquely ergodic*, which means that the invariant measure is unique. As a consequence, for *every point* $(\psi, \theta) \in \mathcal{C}_\psi$ its images $\{\theta + 2n\psi, \quad n \in \mathbb{Z}\}$ are dense and uniformly distributed[1] on $\mathcal{C}_\psi$; this last fact is sometimes referred to as Weyl's theorem [**Pet83**, pp. 49–50].

[1]A sequence of points $x_n \in \mathcal{C}$ on a circle $\mathcal{C}$ is said to be uniformly distributed if for any interval $I \subset \mathcal{C}$ we have $\lim_{N\to\infty} \#\{n\colon 0 < n < N,\ a_n \in I\}/N = \text{length}(I)/\text{length}(\mathcal{C})$.

EXERCISE 1.6. Show that every segment of the particle's trajectory between consecutive collisions is tangent to the smaller circle $S_\psi = \{x^2 + y^2 = \cos^2\psi\}$ concentric to the disk $\mathcal{D}$. Show that if ψ/π is irrational, the trajectory densely fills the ring between $\partial\mathcal{D}$ and the smaller circle S_ψ (see Fig. 1.4).

Remark: One can clearly see in Fig. 1.4 that the particle's trajectory looks denser near the inner boundary of the ring (it "focuses" on the inner circle). If the particle's trajectory were the path of a laser ray and the border of the unit disk were a perfect mirror, then it would feel "very hot" there on the inner circle. For this reason, the inner circle is called a *caustic* (which means "burning" in Greek).

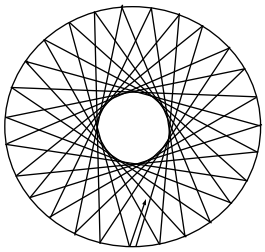

FIGURE 1.4. A nonperiodic trajectory.

EXERCISE 1.7. Can the trajectory of the moving particle be dense in the entire disk $\mathcal{D}$? (Answer: No.)

EXERCISE 1.8. Does the map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ preserve any absolutely continuous invariant measure $d\mu = f(\theta,\psi)\,d\theta\,d\psi$ on $\mathcal{M}$? Answer: Any measure whose density $f(\theta,\psi) = f(\psi)$ is independent of θ is $\mathcal{F}$-invariant.

Next, we can fix the speed of the moving particle due to the following facts.

EXERCISE 1.9. Show that $\|v_t\| = \text{const}$, so that the speed of the particle remains constant at all times.

EXERCISE 1.10. Show that if we change the speed of the particle, say we set $\|v\|_{\text{new}} = c\,\|v\|_{\text{old}}$ with some $c > 0$, then its trajectory will remain unchanged, up to a simple rescaling of time: $q_t^{\text{new}} = q_{ct}^{\text{old}}$ and $v_t^{\text{new}} = v_{ct}^{\text{old}}$ for all $t \in \mathbb{R}$.

Thus, the speed of the particle remains constant and its value is not important. It is customary to set the speed to one: $\|v\| = 1$. Then the velocity vector at time t can be described by an angular coordinate ω_t so that $v_t = (\cos\omega_t, \sin\omega_t)$ and $\omega_t \in [0, 2\pi]$ with the endpoints 0 and 2π being identified.

Now, the collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ represents collisions only. To describe the motion of the particle inside $\mathcal{D}$, let us consider all possible *states* (q, v), where $q \in \mathcal{D}$ is the position and $v \in S^1$ is the velocity vector of the particle. The space of all states (called the *phase space*) is then a three-dimensional manifold $\Omega\colon = \mathcal{D} \times S^1$, which is, of course, a solid torus (doughnut).

The motion of the billiard particle induces a continuous group of transformations of the torus Ω into itself. Precisely, for every $(q, v) \in \Omega$ and every $t \in \mathbb{R}$ the billiard particle starting at (q, v) will come to some point $(q_t, v_t) \in \Omega$ at time t.

Thus we get a map $(q, v) \mapsto (q_t, v_t)$ on Ω, which we denote by Φ^t. The family of maps $\{\Phi^t\}$ is a *group*; i.e. $\Phi^t \circ \Phi^s = \Phi^{t+s}$ for all $t, s \in \mathbb{R}$. This family is called the *billiard flow* on the phase space.

Let us consider a modification of the circular billiard. Denote by $\mathcal{D}_+$ the upper half disk $x^2 + y^2 \le 1$, $y \ge 0$, and let a point particle move inside $\mathcal{D}_+$ and bounce off $\partial\mathcal{D}_+$. (A delicate question arises here: what happens if the particle hits $\partial\mathcal{D}_+$ at $(1, 0)$ or $(-1, 0)$, since there is no tangent line to $\partial\mathcal{D}_+$ at those points? We address this question in the next section.)

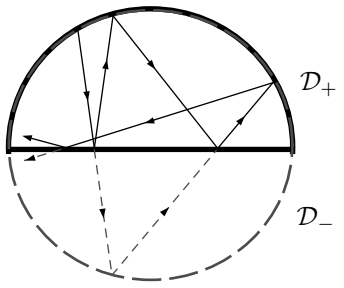

FIGURE 1.5. Billiard in the upper half circle.

A simple trick allows us to reduce this model to a billiard in the full unit disk $\mathcal{D}$. Denote by $\mathcal{D}_-$ the closure of $\mathcal{D} \setminus \mathcal{D}_+$, i.e. the mirror image of $\mathcal{D}_+$ across the x axis $L = \{y = 0\}$. When the particle hits L, its trajectory gets reflected across L, but we will also draw its continuation (mirror image) below L. The latter will evolve in $\mathcal{D}_-$ symmetrically to the real trajectory in $\mathcal{D}_+$ until the latter hits L again. Then these two trajectories will merge and move together in $\mathcal{D}_+$ for a while until the next collision with L, at which time they split again (one goes into $\mathcal{D}_-$ and the other into $\mathcal{D}_+$), etc.

It is important that the second (imaginary) trajectory never actually gets reflected off the line L; it just crosses L every time. Thus it evolves as a billiard trajectory in the full disk $\mathcal{D}$ as described above. The properties of billiard trajectories in $\mathcal{D}_+$ can be easily derived from those discussed above for the full disk $\mathcal{D}$. This type of reduction is quite common in the study of billiards.

EXERCISE 1.11. Prove that periodic trajectories in the half-disk $\mathcal{D}_+$ correspond to periodic trajectories in the full disk $\mathcal{D}$. Note, however, that the period (the number of reflections) may differ.

EXERCISE 1.12. Investigate the billiard motion in a quarter of the unit disk $x^2 + y^2 \le 1$, $x \ge 0$, $y \ge 0$.

1.2. Billiard in a square

Here we describe another simple example – a billiard in the unit square $\mathcal{D} = \{(x, y) \colon 0 \le x, y \le 1\}$; see Fig. 1.6. The laws of motion are the same as before, but this system presents new features.

First of all, when the moving particle hits a vertex of the square $\mathcal{D}$, the reflection rule (1.2) does not apply (there is no normal vector n at a vertex). The particle

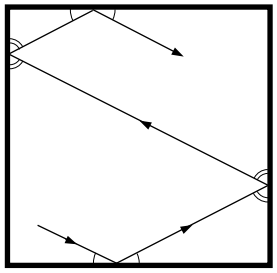

FIGURE 1.6. Billiard in a square.

then stops and its trajectory terminates. We will discuss this exceptional situation later. First we consider regular trajectories that never hit the vertices.

Let $v_t = (u_t, w_t)$ denote the velocity vector of the moving particle at time t (in the x, y coordinates). If it hits a vertical side of $\mathcal{D}$ at time t, then u_t changes sign ($u_{t+0} = -u_{t-0}$) and w_t remains unchanged. If the particle hits a horizontal side of $\mathcal{D}$, then w_t changes sign ($w_{t+0} = -w_{t-0}$) and u_t remains unchanged. Thus,

$$u_t = (-1)^m u_0 \quad \text{and} \quad w_t = (-1)^n w_0, \tag{1.4}$$

where m and n denote the number of collisions with vertical and, respectively, horizontal sides of $\mathcal{D}$ during the time interval $(0, t)$.

EXERCISE 1.13. Show that if $u_0 \neq 0$ and $w_0 \neq 0$ (and assuming the particle never hits a vertex), then all four combinations $(\pm u_0, \pm w_0)$ appear along the particle's trajectory infinitely many times.

Next we make use of the trick shown in Fig. 1.5. Instead of reflecting the trajectory of the billiard particle in a side of $\partial\mathcal{D}$, we reflect the square $\mathcal{D}$ across that side and let the particle move straight into the mirror image of $\mathcal{D}$. If we keep doing this at every collision, our particle will move along a straight line through the multiple copies of $\mathcal{D}$ obtained by successive reflections (the particle "pierces" a chain of squares; see Fig. 1.7). This construction is called the *unfolding* of the billiard trajectory. To recover the original trajectory in $\mathcal{D}$, one *folds* the resulting string of adjacent copies of $\mathcal{D}$ back onto $\mathcal{D}$.

We denote the copies of $\mathcal{D}$ by

$$\mathcal{D}_{m,n} = \{(x, y)\colon\ m \leq x \leq m+1,\ n \leq y \leq n+1\}. \tag{1.5}$$

EXERCISE 1.14. Show that if m and n are even, then the folding procedure transforms $\mathcal{D}_{m,n}$ back onto $\mathcal{D} = \mathcal{D}_{0,0}$ by translations $x \mapsto x - m$ and $y \mapsto y - n$, thus preserving orientation of both x and y. If m is odd, then the orientation of x is reversed (precisely, $x \mapsto m + 1 - x$). If n is odd, then the orientation of y is reversed (precisely, $y \mapsto n + 1 - y$). Observe that these rules do not depend on the particular trajectory that was originally unfolded.

The squares $\mathcal{D}_{m,n}$ with $m, n \in \mathbb{Z}$ tile, like blocks, the entire plane $\mathbb{R}^2$. Any regular billiard trajectory unfolds into a directed straight line on the plane, and any directed line (which avoids the sites of the integer lattice) folds back into a billiard trajectory. A trajectory hits a vertex of $\mathcal{D}$ iff the corresponding line runs into a site of the integer lattice.

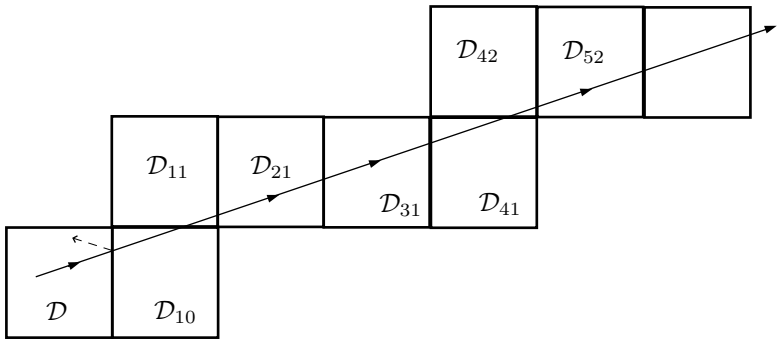

FIGURE 1.7. Unfolding a billiard trajectory.

The structure of blocks $\mathcal{D}_{m,n}$ with the respective folding rules is clearly periodic, in which the 2×2 square

$$\mathbb{K}_2 = \{(x, y) \colon 0 \le x, y \le 2\}$$

plays the role of a fundamental domain – the entire plane is covered by parallel translations of $\mathbb{K}_2$. Thus the standard projection of $\mathbb{R}^2$ onto $\mathbb{K}_2$ transforms unfolded trajectories into directed straight lines on the 2×2 torus Tor^2 (the latter is obtained by identifying the opposite sides of the square $\mathbb{K}_2$). The billiard in the unit square $\mathcal{D}$ thus reduces to the simple linear flow on a flat 2×2 torus Tor^2, in which points move with constant (unit) velocity vectors.

The linear flow on a flat torus is one of the standard examples in ergodic theory; cf. Appendix C and [**KH95, Pet83, Sin76**]. Its main properties are these:

- if a trajectory has rational slope $dy/dx \in \mathbb{Q}$, then it is periodic (it runs along a closed geodesic);
- if a trajectory has irrational slope $dy/dx \notin \mathbb{Q}$, then it is dense (its closure is the whole torus).

This translates into the following alternative for regular billiard trajectories in the unit square $\mathcal{D}$:

COROLLARY 1.15. *If $w_0/u_0 \in \mathbb{Q}$, then the corresponding regular billiard trajectory in the unit square $\mathcal{D}$ is periodic. If $w_0/u_0 \notin \mathbb{Q}$, then the corresponding regular billiard trajectory is dense.*

EXERCISE 1.16. Extend this result to the billiard in a rectangle $\mathcal{R}$ with sides a and b. Answer: a regular billiard trajectory in $\mathcal{R}$ is periodic iff $(aw_0)/(bu_0) \in \mathbb{Q}$; otherwise it is dense. Hint: Transform the rectangle into the unit square by scaling the coordinates: $(x, y) \mapsto (x/a, y/b)$. Argue that the billiard trajectories in $\mathcal{R}$ will be thus transformed into those in $\mathcal{D}$.

EXERCISE 1.17. Extend the above result to billiards in the following polygons: an equilateral triangle, a right isosceles triangle, a right triangle with the acute angle $\pi/6$, and a regular hexagon. What is common about these polygons? (Note that the billiard in a hexagon does not reduce to a geodesic flow on a torus. Does it reduce to a geodesic flow on another manifold?)

The phase space of the billiard system in the unit square $\mathcal{D}$ is the three-dimensional manifold $\Omega = \mathcal{D} \times S^1$; cf. the previous section. The billiard flow Φ^t is defined for all times $-\infty < t < \infty$ on regular trajectories. On exceptional trajectories (which hit a vertex of $\mathcal{D}$ at some time), the flow is defined only until the trajectory terminates in a vertex.

EXERCISE 1.18. Show that the set of exceptional trajectories is a countable union of 2D surfaces in Ω.

We see that the set of exceptional trajectories is negligible in the topological and measure-theoretic sense (it has zero Lebesgue measure and is an F_σ set, i.e. a countable union of nowhere dense closed subsets), but still its presence is bothersome. For the billiard in a square, though, one can get rid of them altogether by extending the billiard flow by continuity.

EXERCISE 1.19. Show that the flow Φ^t can be uniquely extended by continuity to all exceptional trajectories. In that case every trajectory hitting a vertex of $\mathcal{D}$ will simply reverse its course and run straight back; see Fig. 1.8.

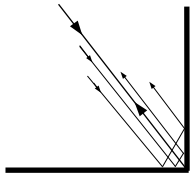

FIGURE 1.8. Extension of the flow near a vertex.

The above extension defines the billiard flow Φ^t on the entire phase space Ω and makes it continuous everywhere. We will assume this extension in what follows. We remark, however, that in generic billiards such nice extensions are rarely possible; see Section 2.8.

Now, the action of the flow Φ^t on the phase space Ω can be fully described as follows. For every unit vector $v_0 = (u_0, w_0) \in S^1$, consider the set

$$\mathcal{L}_{v_0} = \{(q, v) \in \Omega\colon\ q \in \mathcal{D},\ v = (\pm u_0, \pm w_0)\}$$

(the two signs are, of course, independent). Due to (1.4), each set $\mathcal{L}_{v_0}$ remains invariant under the flow Φ^t.

Suppose first that $u_0 \neq 0$ and $v_0 \neq 0$; then $\mathcal{L}_{v_0}$ is the union of four squares obtained by "slicing" Ω at the four "levels" corresponding to the vectors $(\pm u_0, \pm w_0)$; see Fig. 1.9.

EXERCISE 1.20. Check that the four squares constituting the set $\mathcal{L}_{v_0}$ can be glued together along their boundaries, and obtain a smooth closed surface without boundary (a 2×2 torus) $\mathbb{T}^2_{v_0}$ on which the billiard flow will coincide with the linear flow along the vector v_0 (i.e. the flow on $\mathbb{T}^2_{v_0}$ will be defined by differential equations $\dot{x} = u_0$, $\dot{y} = w_0$). Hint: The assembly of the torus $\mathbb{T}^2_{v_0}$ from the squares of $\mathcal{L}_{v_0}$ is very similar to the reduction of the billiard dynamics in $\mathcal{D}$ to the geodesic flow on the 2×2 torus described above (in fact, these two procedures are equivalent).

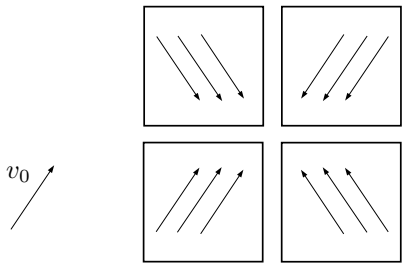

FIGURE 1.9. Four squares constituting $\mathcal{L}_{v_0}$.

Now, it is a standard fact in ergodic theory (cf. Appendix C) that the linear flow on a 2D torus defined by $\dot{x} = u_0$, $\dot{y} = w_0$ is periodic if $w_0/u_0 \in \mathbb{Q}$ and ergodic (furthermore, uniquely ergodic) if $w_0/u_0 \notin \mathbb{Q}$. In the latter case every trajectory is dense and uniformly distributed[2] on the torus.

In the two remaining cases (first $u_0 = 0$, and second $w_0 = 0$) the set $\mathcal{L}_{v_0}$ consists of just two squares. We leave their analysis to the reader as an easy exercise.

This fully describes the action of the flow $\Phi^t \colon \Omega \to \Omega$ for the billiard in the unit square.

1.3. A simple mechanical model

As a motivation for the study of billiards, one usually describes a simple model of two moving particles in a one-dimensional container. It reduces to a billiard in a right triangle, which is similar to a billiard in a square. We describe this model here; see also [**CFS82, CM03**].

Consider a system of two point particles of masses m_1 and m_2 on a unit interval $0 \le x \le 1$. The particles move freely and collide elastically with each other and with the 'walls' at $x = 0$ and $x = 1$. Let x_1 and x_2 denote the positions of the particles and u_1 and u_2 their velocities. Since the particles collide upon contact, their positions remain ordered; we assume that $x_1 \le x_2$ at all times.

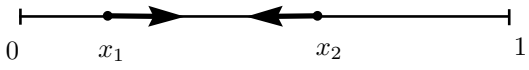

FIGURE 1.10. Two particles in a unit interval.

Next we describe collisions. When a particle hits a wall, it simply reverses its velocity. When the two particles collide with each other, we denote by u_i^- the precollisional velocity and by u_i^+ the postcollisional velocity of the ith particle, $i = 1, 2$. The law of elastic collisions requires the conservation of the total momentum, i.e.

$$m_1 u_1^+ + m_2 u_2^+ = m_1 u_1^- + m_2 u_2^-$$

[2] A line x_t on a 2D torus Tor^2 is said to be uniformly distributed if for any rectangle $R \subset \mathrm{Tor}^2$ we have $\lim_{T\to\infty} \mathbf{m}\big(\{t \colon 0 < t < T,\ x_t \in R\}\big)/T = \text{ area}(R)/\text{ area}(\mathrm{Tor}^2)$. Here $\mathbf{m}$ is the Lebesgue measure on $\mathbb{R}$.

and the total kinetic energy, i.e.

$$m_1[u_1^+]^2 + m_2[u_2^+]^2 = m_1[u_1^-]^2 + m_2[u_2^-]^2. \tag{1.6}$$

Solving these equations gives

$$u_1^+ = u_1^- + \frac{2m_2}{m_1+m_2}(u_2^- - u_1^-)$$

and

$$u_2^+ = u_2^- + \frac{2m_1}{m_1+m_2}(u_1^- - u_2^-)$$

(we recommend the reader derive these formulas for an exercise). Note that if $m_1 = m_2$, then the particles simply exchange their velocities: $u_1^+ = u_2^-$ and $u_2^+ = u_1^-$.

The variables x_i and u_i are actually inconvenient, so we will work with new variables defined by

$$q_i = x_i\sqrt{m_i} \qquad \text{and} \qquad v_i = dq_i/dt = u_i\sqrt{m_i} \tag{1.7}$$

for $i = 1, 2$. Now the positions of the particles are described by a point $\mathbf{q} = (q_1, q_2) \in \mathbb{R}^2$ (it is called a *configuration point*). The set of all configuration points (called the *configuration space*) is the right triangle

$$\mathcal{D} = \{\mathbf{q} = (q_1, q_2) : 0 \le q_1/\sqrt{m_1} \le q_2/\sqrt{m_2} \le 1\}.$$

The velocities of the particles are described by the vector $\mathbf{v} = (v_1, v_2)$. Note that the energy conservation law (1.6) implies that $\|\mathbf{v}\| = \text{const}$; thus we can set $\|\mathbf{v}\| = 1$.

The state of the system is described by a pair $(\mathbf{q}, \mathbf{v})$. The configuration point $\mathbf{q}$ moves in $\mathcal{D}$ with velocity vector $\mathbf{v}$. When the first particle collides with the wall ($x_1 = 0$), the configuration point hits the left side $q_1 = 0$ of the triangle $\mathcal{D}$. When the second particle collides with the wall ($x_2 = 1$), the point $\mathbf{q}$ hits the upper side $q_2/\sqrt{m_2} = 1$ of $\mathcal{D}$. When the particles collide with each other, the point $\mathbf{q}$ hits the hypotenuse $q_1/\sqrt{m_1} = q_2/\sqrt{m_2}$ of $\mathcal{D}$.

Exercise 1.21. Prove that the velocity vector $\mathbf{v}$ changes at collisions so that it gets reflected at $\partial\mathcal{D}$ according to the law 'the angle of incidence is equal to the angle of reflection'.

Thus, the motion of the configuration point $\mathbf{q}$ is governed by the billiard rules. Hence the evolution of the mechanical model of two particles in a unit interval reduces to billiard dynamics in a right triangle.

If $m_1 = m_2$, we obtain a billiard in a right isosceles triangle, which readily reduces to a billiard in a square; see Exercise 1.17. For generic mass ratio m_1/m_2, we obtain a billiard in a generic right triangle, which may be rather complicated (such billiards are not covered in our book).

One complication arises when a billiard trajectory hits a corner point of $\mathcal{D}$. Hitting the vertex of the right angle corresponds to an event when both particles simultaneously collide with opposite walls. Then their further motion is clearly well defined; thus the billiard trajectory can easily be continued (cf. Exercise 1.19).

However, hitting the vertex of an acute angle of $\mathcal{D}$ corresponds to an event when both particles simultaneously collide with the *same* wall (either $x = 0$ or $x = 1$). In this case, for generic m_1 and m_2, the billiard flow cannot be extended by continuity, as nearby trajectories hitting the two adjacent sides in different order will come back to $\mathcal{D}$ along different lines; see Fig. 1.11.

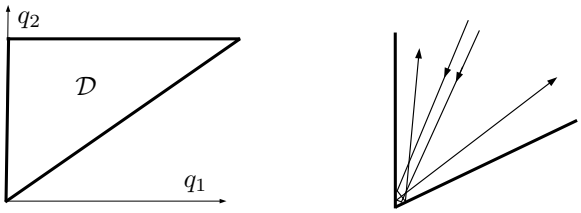

FIGURE 1.11. The right triangle $\mathcal{D}$; hitting the vertex of an acute angle.

In mechanical terms, hitting the vertex of an acute angle of $\mathcal{D}$ corresponds to a *multiple collision*. Such exceptional events usually cannot be resolved by the laws of classical mechanics.

1.4. Billiard in an ellipse

We proceed to yet another simple example that admits a completely elementary analysis – the billiard in an ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

with some $a > b > 0$. In fact, it was this example that Birkhoff described in the very first book on mathematical billiards in 1927 [**Bi27**, Chapter VIII].

We denote by $\mathcal{D}$ the domain bounded by the ellipse (it will be our billiard table). Let F_1 and F_2 denote the foci of the ellipse, and observe that they lie on the x axis. The ellipse is the locus of points $A \in \mathbb{R}^2$ such that

$$\operatorname{dist}(A, F_1) + \operatorname{dist}(A, F_2) = \text{ const.}$$

EXERCISE 1.22. Let $A \in \partial\mathcal{D}$ and L denote the tangent line to the ellipse at A. Prove that the segments AF_1 and AF_2 make equal angles with L. (This fact is known in projective geometry as the Poncelet theorem.) Hint: Reflect the point F_2 across the tangent line L and show that its image will lie on the line AF_1.

Thus, if a billiard trajectory passes through one focus, then it reflects at a point $A \in \partial\mathcal{D}$ on the ellipse and runs straight into the other focus. Such a trajectory will then pass through a focus after every reflection; see Fig. 1.12.

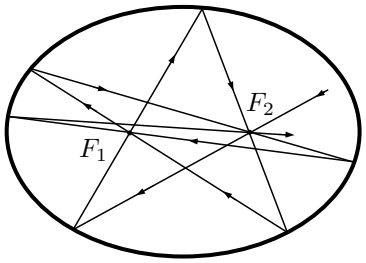

FIGURE 1.12. A trajectory passing through the foci.

EXERCISE 1.23. Show that every trajectory passing through the foci F_1 and F_2 converges to the major axis of the ellipse (the x axis).

By the way, the major and the minor axes of the ellipse are clearly two periodic trajectories – they run back and forth between their endpoints.

In Section 1.1 we used the coordinates ψ and θ to describe collisions in a circular billiard, and the cyclic coordinate θ was actually the arc length parameter on the circle (Remark 1.2). Here we use two coordinates ψ and r, where ψ is the same angle of reflection as in Section 1.1 and r is an arclength parameter on the ellipse. We choose the reference point $r = 0$ as the rightmost point $(a, 0)$ on the ellipse and orient r counterclockwise. Note that $0 \le r \le |\partial\mathcal{D}|$ and $0 \le \psi \le \pi$.

The collision space $\mathcal{M}$ is again a cylinder whose base is the ellipse and whose height is π. It is shown in Fig. 1.13 as a rectangle $[0, |\partial\mathcal{D}|] \times [0, \pi]$, but we keep in mind that the left and right sides of this rectangle must be identified. The motion of the billiard particle, from collision to collision, induces the collision map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$.

EXERCISE 1.24. Verify that the trajectories passing through the foci lie on a closed curve on the surface $\mathcal{M}$. Determine its shape. Answer: It is the ∞-shaped curve in Fig. 1.13 that separates the white and grey areas.

Thus, the trajectories passing through the foci make a special (one-dimensional) family in $\mathcal{M}$.

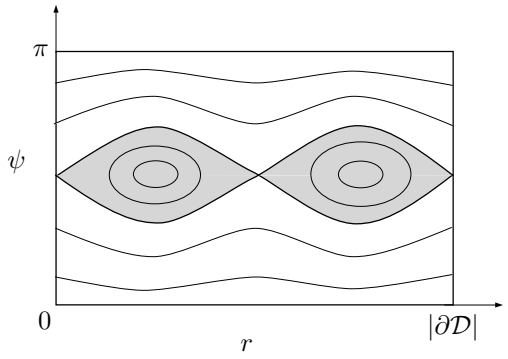

FIGURE 1.13. The collision space of an elliptic billiard.

EXERCISE 1.25. Show that if the trajectory of the billiard particle crosses the segment F_1F_2 joining the foci, then it reflects at $\partial\mathcal{D}$ and crosses this segment again. Similarly, if the trajectory crosses the major axis beyond the segment F_1F_2, say to the left of it, then after one or more reflections at $\partial\mathcal{D}$ it will cross the major axis to the right of this segment, etc.

The previous exercise shows that there are trajectories of two types: those crossing the inner segment F_1F_2 of the major axis after every reflection (we call them *inner* trajectories) and those going around this segment (we call them *outer* trajectories).

EXERCISE 1.26. Verify that the inner trajectories fill the white area in Fig. 1.13, and outer trajectories fill the grey area.

The following is the most important property of elliptic billiards:

THEOREM 1.27. *For every outer trajectory there is an ellipse with foci F_1 and F_2 that is tangent to each link of that trajectory. For every inner trajectory there is a hyperbola with foci F_1 and F_2 that is tangent to each link (or its linear extension) of that trajectory.*

PROOF. We prove only the first statement (about the outer trajectories); the proof of the second is similar. The argument is pretty elementary and illustrated in Fig. 1.14. Here A_1A and A_2A are two successive links of an outer trajectory. The points B_1 and B_2 are obtained by reflecting the foci F_1 and F_2 across the lines A_1A and A_2A, respectively. The four angles $\angle B_1AA_1$, $\angle A_1AF_1$, $\angle F_2AA_2$ and $\angle A_2AB_2$ are equal. Hence the triangles AB_1F_2 and AB_2F_1 are congruent, in particular $|B_1F_2| = |B_2F_1|$. Therefore

$$|F_1C_1| + |F_2C_1| = |F_1C_2| + |F_2C_2|,$$

where C_1 and C_2 are the points of intersection of A_1A with B_1F_2 and A_2A with B_2F_1, respectively. Thus, the points C_1 and C_2 belong to the same ellipse with foci F_1 and F_2, and the lines A_1A and A_2A are tangent to that ellipse. □

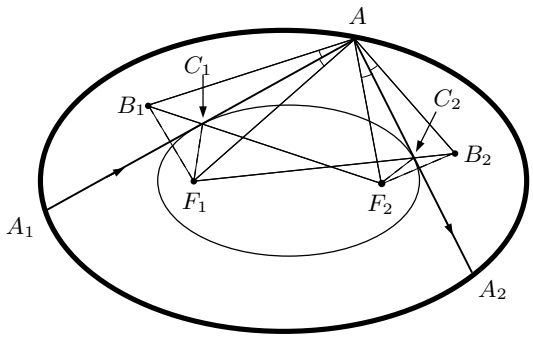

FIGURE 1.14. Proof of Theorem 1.27.

If every link of a billiard trajectory is tangent to a certain given curve, then that curve is called a *caustic*. Fig. 1.15 shows an elliptic caustic for an outer trajectory and a hyperbolic caustic for an inner trajectory. The term 'caustic' is borrowed from optics, where it means a curve on which light rays focus after being reflected off a mirror (we have seen caustics in circular billiards in Section 1.1). Fig. 1.15 demonstrates the concentration of rays on caustics (compare it to Fig. 1.4).

All the trajectories tangent to one elliptic caustic lie on a closed curve in the collision space $\mathcal{M}$. Such curves are shown as 'horizontal waves' in the white area in Fig. 1.13 (remember that the left and right sides of the rectangle need to be identified). Every such curve is obviously invariant under the map $\mathcal{F}$.

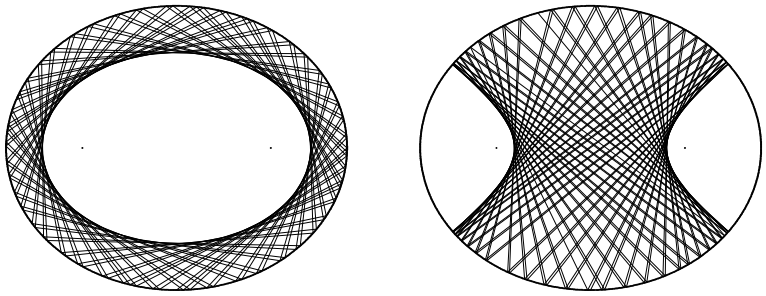

FIGURE 1.15. Elliptic and hyperbolic caustics in the elliptic billiard.

EXERCISE 1.28. On each invariant curve the map $\mathcal{F}$ is conjugate to a rigid circle rotation through some angle (that angle is called the *rotation number*). Show that the rotation number changes continuously and monotonically with the invariant curve. Hint: Consider two outer trajectories starting at the same point $A_0 \in \partial\mathcal{D}$ but with distinct elliptical caustics. Denote by A'_n the reflection points of the trajectory whose elliptical caustic is smaller and by A''_n those of the other trajectory. Observe that the sequence $\{A'_n\}$ will move along the ellipse faster than $\{A''_n\}$ does; see Fig. 1.16.

The action of the map $\mathcal{F}$ on each invariant curve can be analyzed explicitly and the rotation number can be computed analytically (see [**Be01**, Sections 2.5 and 3.2]) but we will not go that far.

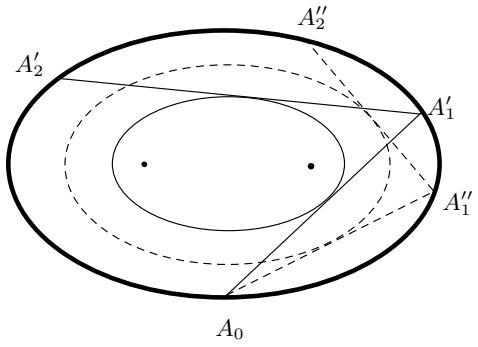

FIGURE 1.16. Exercise 1.28.

Next, all the trajectories tangent to one hyperbolic caustic lie on two closed curves in $\mathcal{M}$, one inside each half of the ∞-shaped grey domain. Such curves appear as ovals in Fig. 1.13. The map $\mathcal{F}$ transforms each oval into an identical oval within the other half of the ∞-shaped grey domain. Thus the union of the two identical (symmetric) ovals will be invariant under $\mathcal{F}$, and each oval separately will be invariant under $\mathcal{F}^2$.

Therefore, the collision space $\mathcal{M}$ of an elliptical billiard is completely foliated by invariant curves. In this sense, the elliptical billiard is similar to those in a circle and in a square. In physics, such models belong to a special class: if the phase

space of a system is foliated by one-dimensional invariant submanifolds, the system is said to be *integrable*; the dynamics in such a system is completely regular. Thus, billiards in circles, squares and ellipses are completely regular.

1.5. A chaotic billiard: pinball machine

The simple examples were given in the previous sections for the sake of introducing some basic features of billiards to the novice reader. But they should not be taken as typical; in fact their dynamical characteristics are quite special and in a sense opposite to those of chaotic billiards which will be covered in the rest of the book. Here we will take a glimpse at something that happens in chaotic billiards.

Imagine you are playing a pinball machine. A small ball shoots from a cannon in the right bottom corner of a rectangular table; then it bounces off the edges until it either hits the target (then you win) or falls through an opening in the bottom (goes down the drain; then you lose). The target might be a special figure on the table that registers the hit when the ball touches it. To prevent direct hits, assume the target is screened from the cannon and hitting the screen is forbidden by the rules. Then the ball has to bounce off the edges before reaching the target; see Fig. 1.17. It is quite an unusual pinball machine, but for us it is a good starting example.

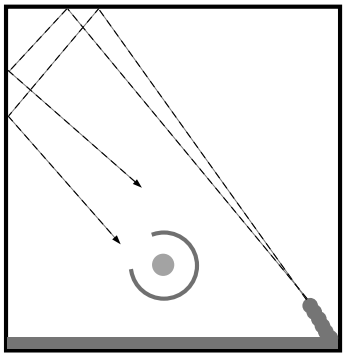

FIGURE 1.17. A pinball machine. The target is the grey disk screened from the cannon by a dark grey arc.

Suppose you can rotate the cannon to change the angle at which the ball shoots out. After you miss once, you can adjust the shooting angle and send the ball more accurately into the target. This is a relatively easy task (illustrated in Fig. 1.17), as the trajectory of the ball (in a rectangular billiard) is very simple and predictable. Also, you do not need to aim with absolute precision.

EXERCISE 1.29. Suppose the target is a disk of radius r and the moving ball is a point particle. Let L denote the distance covered by the ball from the cannon to the target. Show that if the shooting angle is off by less than r/L (radians), then the ball still hits the target.

Now let us make the task more realistic and challenging by installing some bumpers (round pillars) all over the table (see Fig. 1.18) so that our moving ball

will bounce between the bumpers on its way to the target (or down the drain). Anyone who has played real pinball machines can easily imagine such a process.

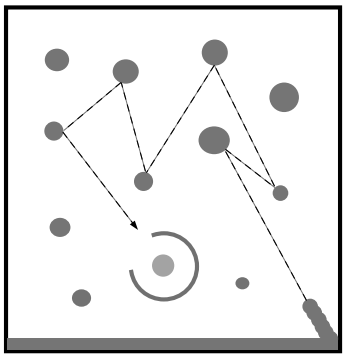

FIGURE 1.18. A pinball machine with bumpers (dark grey disks).

Would it be easy to adjust the cannon on this new table? Obviously not. The route of the ball is complicated and almost unpredictable, as it may bounce off different bumpers. Even finding the right sequence of bumpers that the ball needs to hit before it reaches the target (and avoids the screen) is not a simple task. Professional billiard players solve a similar problem when trying the send a ball to a pocket, so that it hits one or more other balls.

Furthermore, the cannon must be aimed with almost ultimate precision, as a tiny error in the shooting angle may send the ball rolling down along a completely wrong path. This is illustrated in Fig. 1.19, where just two successive bounces are shown. It is quite clear that the instability of the ball's motion increases with every subsequent reflection off a bumper. Again, professional billiard players know that if their ball needs to hit more than one other ball before sending one of them into a pocket, their task is very difficult. Furthermore, if the ball must hit three or more other balls, the task is almost impossible.

A rectangular table with round bumpers installed on it is a classical example of a chaotic billiard. The motion of the billiard particle on such a table is complicated and unpredictable. To the naked eye, it may look like a wild dance between the walls of the table, without any pattern or logic (that is why pinball machines are so attractive!). The lack of predictability is characteristic of chaotic billiards.

Furthermore, slight changes in the initial position and/or velocity of the particle quickly lead to large deviations (such as that in Fig. 1.19), so after just a few collisions with bumpers two trajectories, initially very close together, will separate and move far apart from each other, as if they were unrelated. This instability (also known as *sensitivity to initial conditions*) is another characteristic feature of chaotic billiards (and chaotic dynamics in general).

In practical terms, the best thing the player can do in our game is to shoot randomly and watch the ball running all over the table, bouncing around between bumpers – there will always be a chance that it hits the target 'by accident'. This is essentially a game of chance, just like flipping a coin, rolling a die, or playing cards. We will see in Chapters 6 and 7 that the motion in a chaotic billiard is indeed essentially random and is best described in terms of probability theory.

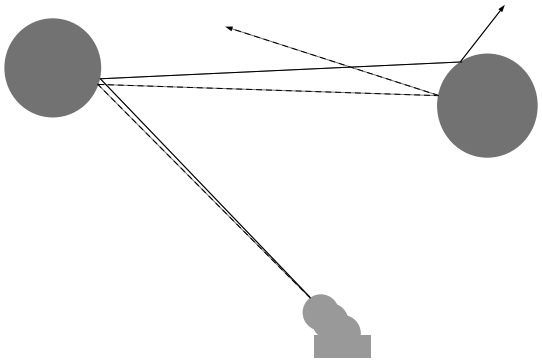

FIGURE 1.19. A ball bouncing off two bumpers: a slight error in the initial shooting angle results in a dramatic deflection in the end.

Our toy example actually has a lot in common with a classical model of statistical physics, called the Lorentz gas. In that model, a small ball (electron) bounces between large fixed disks (molecules) that make a regular periodic (crystalline) structure. We will present it in Chapter 5.6.

We will not attempt to go beyond this very informal introduction to the realm of chaotic billiards, leaving all formalities till future chapters. Interested readers may find a more extensive description of chaotic billiards, including computer illustrations, in [**Be01**, Section 1.1].

CHAPTER 2

Basic constructions

Billiard dynamics in planar regions is commonly defined as follows:

DEFINITION 2.1. 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 $\partial\mathcal{D}$.

This definition is clear and easy-to-use if one has a particular domain $\mathcal{D}$ at hand. In general studies, however, one needs to specify the class of domains to be considered. One has to be especially careful about the boundary $\partial\mathcal{D}$ – whether or not to allow $\partial\mathcal{D}$ to have infinite length, or unbounded curvature, or infinitely many inflection points, etc.

In most research articles on billiards, it is assumed (explicitly or implicitly) that these and other pathologies do not occur. A notable exception is Strelcyn's paper [**St86**], which covers very general billiards, but it clearly demonstrates that irregularities of the boundary may lead to serious complications in the analysis of the dynamics and render its properties barely tractable.

Here we exclude many unpleasant pathologies from our studies by making appropriate assumptions on the domain $\mathcal{D}$ (see Assumptions A1–A3 in Section 2.1 and Assumption A4 in Section 2.4). This allows us to proceed smoothly and avoid many troubles so characteristic of more general studies [**St86**]. Still, our assumptions A1–A4 are general enough to cover practically all presently known chaotic billiards.

2.1. Billiard tables

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

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

$$\partial\mathcal{D} = \Gamma = \Gamma_1 \cup \cdots \cup \Gamma_r. \tag{2.1}$$

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

We call $\mathcal{D}$ a *billiard table* and $\Gamma_1, \dots, \Gamma_r$ *walls* or *components* of $\partial\mathcal{D}$.

REMARK 2.2. If $f_i(a_i) \neq f_i(b_i)$, then we call Γ_i an *arc* and denote $\partial\Gamma_i = \{f_i(a_i), f_i(b_i)\}$. If $f_i(a_i) = f_i(b_i)$, then Γ_i is a *closed curve*, which may or may not be entirely C^ℓ smooth. If it is, then f_i can be defined on a circle S^1 rather than $[a_i, b_i]$, and we set $\partial\Gamma_i = \emptyset$. It may happen that a closed curve Γ_i lacks smoothness at the point $f_i(a_i) = f_i(b_i)$; then we denote that point by $\partial\Gamma_i$.

ASSUMPTION A2. The boundary components Γ_i can intersect each other only at their endpoints; i.e.

$$\Gamma_i \cap \Gamma_j \subset \partial\Gamma_i \cup \partial\Gamma_j \qquad \text{for } i \neq j. \tag{2.2}$$

We put

$$\Gamma_* = \partial\Gamma_1 \cup \cdots \cup \partial\Gamma_r \quad \text{and} \quad \tilde{\Gamma} = \Gamma \setminus \Gamma_*. \tag{2.3}$$

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

EXERCISE 2.3. Show that every regular boundary point $x \in \tilde{\Gamma}$ has an open neighborhood $U(x)$ that intersects only one wall Γ_i and is divided by Γ_i into two open connected parts – one lies in the interior of $\mathcal{D}$ and the other in $\mathbb{R}^2 \setminus \mathcal{D}$.

EXERCISE 2.4. Show that every corner point $x \in \Gamma_*$ has an open neighborhood $U(x)$ such that $U(x) \cap \Gamma$ is a disjoint union of $2m$ curves for which x is a common endpoint (here $m = m_x \geq 1$). They divide $U(x)$ into $2m$ open connected domains so that m of them lie in the interior of $\mathcal{D}$ and the other m in $\mathbb{R}^2 \setminus \mathcal{D}$, and these two types of domains alternate as one traverses $U(x)$ around x; see Fig. 2.1.

We call x a *simple corner point* if $m_x = 1$. In our book we mostly deal with simple corner points.

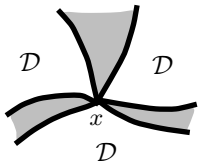

FIGURE 2.1. A corner point x with $m_x = 3$.

In the above exercise, the connected components of $U(x)$ that lie inside $\mathcal{D}$ are called *corners* of the domain $\mathcal{D}$. Each corner is bounded by two curves $\Gamma_i, \Gamma_j \subset \Gamma$ converging to the point x (we call these curves the *sides* of the corner) and characterized by the interior angle γ made by the corresponding tangent lines at x (the angle seen from inside $\mathcal{D}$). If $\gamma = 0$, we call the corner a *cusp*.

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

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

Note that $f_i'' \perp f_i'$, because $\|f_i'\| = \text{const}$. Hence, if $f_i'' \neq 0$, then the pair of vectors f_i', f_i'' is either left or right along the entire curve Γ_i. Accordingly, we distinguish three types of walls:

Flat walls: such that $f'' \equiv 0$;

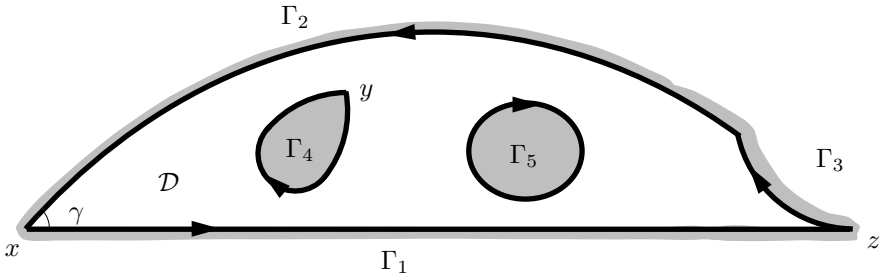

FIGURE 2.2. A billiard table. Here Γ_1 is flat; Γ_2 is focusing; Γ_3 is dispersing; Γ_4 is a closed wall with one corner point, y; Γ_5 is a smooth closed wall; the corner at x has a positive interior angle $\gamma > 0$; and the corner at z is a cusp. The adopted orientation of Γ is shown by arrows.

Focusing walls: such that $f'' \neq 0$ is pointing inside $\mathcal{D}$;
Dispersing walls: such that $f'' \neq 0$ is pointing outside $\mathcal{D}$.

The terms "focusing" and "dispersing" are standard in billiard literature. They refer to the effect the corresponding mirror walls have on reflecting light rays – type (b) walls tend to focus the rays, while type (c) walls tend to disperse them; see Figure 2.3. Focusing walls are also referred to as *convex* and dispersing as *concave*.

We now define the (*signed*) *curvature* on each Γ_i as follows:

$$\mathcal{K} = \begin{cases} 0 & \text{if } \Gamma_i \text{ is flat;} \\ -\|f''\| & \text{if } \Gamma_i \text{ is focusing;} \\ \|f''\| & \text{if } \Gamma_i \text{ is dispersing.} \end{cases} \tag{2.4}$$

Our choice of signs is dictated by a tradition in billiard literature first adopted by Ya. Sinai and his school. However, the opposite choice is sometimes made, too.

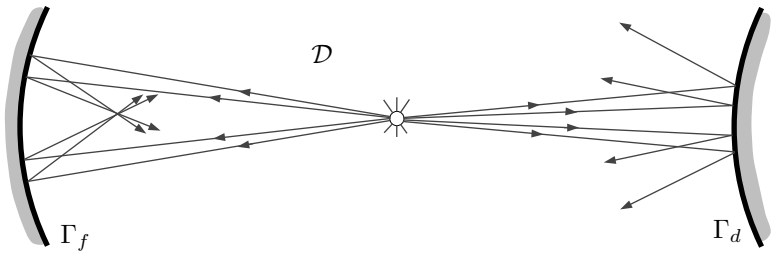

FIGURE 2.3. Reflecting light rays tend to focus near a *focusing* wall Γ_f and disperse near a *dispersing* wall Γ_d.

Accordingly, we define

$$\Gamma_0 = \bigcup_{\mathcal{K}=0} \Gamma_i, \quad \Gamma_- = \bigcup_{\mathcal{K}<0} \Gamma_i, \quad \Gamma_+ = \bigcup_{\mathcal{K}>0} \Gamma_i. \tag{2.5}$$

We note that the curvature of each focusing or dispersing wall Γ_i is bounded away from zero and infinity, due to Assumption A3 and the closedness of Γ_i.

Clearly, each wall Γ_i has finite length. We denote it by $|\Gamma_i|$ and set $|\Gamma| = \sum_i |\Gamma_i|$ to be the total perimeter of $\mathcal{D}$.

2.2. Unbounded billiard tables

We have assumed so far that $\mathcal{D}$ was bounded. Sometimes one deals with unbounded billiard tables $\mathcal{D} \subset \mathbb{R}^2$, and we want to include them in our studies, too. In that case it is reasonable to require that boundary $\partial\mathcal{D}$ be *locally piece-wise smooth*; that is for any large square $\mathbb{K}_B = \{(x_1, x_2) \in \mathbb{R}^2 \colon |x_1| \le B,\ |x_2| \le B\}$ the intersection $\mathcal{D} \cap \mathbb{K}_B$ must have a finitely piece-wise smooth boundary satisfying Assumptions A1–A3.

In many studies of unbounded tables $\mathcal{D}$ one assumes that they have periodic structure. This happens, for example, when there are two orthogonal unit vectors $u, v \in \mathbb{R}^2$ such that

$$q \in \mathcal{D} \iff q + u \in \mathcal{D} \iff q + v \in \mathcal{D}. \tag{2.6}$$

In this case we can choose the coordinate axes parallel to the vectors u and v, so that $\mathbb{R}^2$ can be regarded as the universal cover of the corresponding unit torus $\mathrm{Tor}^2 = \mathbb{R}^2/\mathbb{Z}^2$. Then an unbounded periodic domain $\mathcal{D}$ satisfying (2.6) can be represented by its projection onto Tor^2.

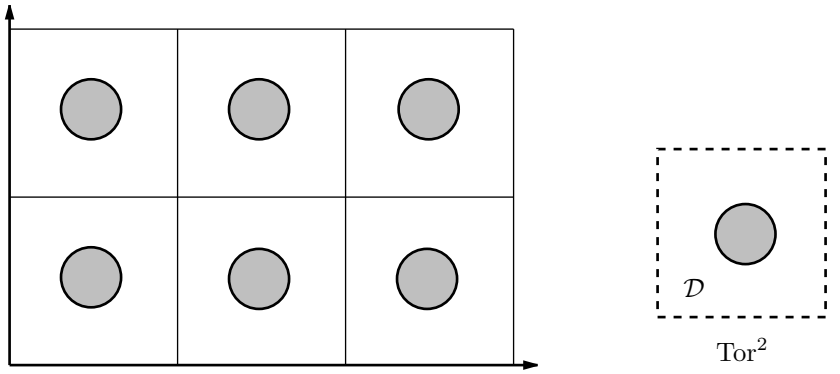

FIGURE 2.4. An unbounded periodic table and its projection on Tor^2.

Slightly abusing our notation, we also denote the above projection of $\mathcal{D}$ onto Tor^2 by $\mathcal{D}$. Obviously, such a domain $\mathcal{D} \subset \mathrm{Tor}^2$ will satisfy Assumptions A1–A3. In addition, we will always assume that $\mathcal{D} \neq \mathrm{Tor}^2$ to avoid trivial "billiards" without walls (or collisions).

We summarize our constructions as follows:

DEFINITION 2.5. A *billiard table* $\mathcal{D}$ is the closure of a bounded open connected domain $\mathcal{D} \subset \mathbb{R}^2$ or $\mathcal{D} \subsetneq \mathrm{Tor}^2$ such that $\partial\mathcal{D}$ satisfies Assumptions A1–A3.

2.3. Billiard flow

Here we construct dynamics on a billiard table. This is not going to be a simple task, since there will be several cases where the construction fails and the trajectory of a billiard particle cannot be defined.

Let $q \in \mathcal{D}$ denote the position of the moving particle and $v \in \mathbb{R}^2$ its velocity vector. Of course, $q = q(t)$ and $v = v(t)$ are functions of time $t \in \mathbb{R}$. When the particle moves inside the table, so that $q \in \operatorname{int} \mathcal{D}$, it maintains a constant velocity:

$$\dot{q} = v \qquad \text{and} \qquad \dot{v} = 0 \tag{2.7}$$

(here the dot denotes the time derivative).

When the particle collides with the regular part of the boundary, i.e. $q \in \tilde{\Gamma}$, its velocity vector instantaneously gets reflected across the tangent line to Γ at the point q. This is specified by the classical rule "the angle of incidence is equal to the angle of reflection" (see Fig. 2.5), and it can be expressed by

$$v^+ = v^- - 2\,\langle v, n\rangle\, n, \tag{2.8}$$

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.

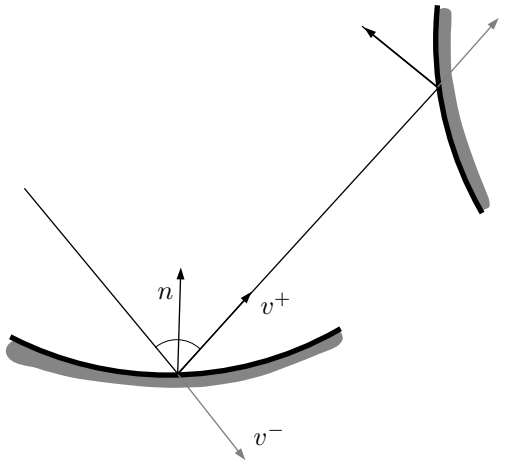

FIGURE 2.5. Velocity vector changes at collisions.

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. This is one of the complications that make the analysis of billiard dynamics difficult (we will see more later).

The equations of motion (2.7)–(2.8) preserve the norm $\|v\|$, and it is customary to set it to one: $\|v\| = 1$.

DEFINITION 2.6. A collision is said to be *regular* if $q \in \tilde{\Gamma}$ and the vector v^- is not tangent to Γ. In this case $v^+ \neq v^-$ (obviously, v^- must be pointing outside $\mathcal{D}$, and v^+ inside $\mathcal{D}$). If v^- is tangent to Γ at the point of collision, then $v^+ = v^-$, and such a collision is said to be *grazing* or *tangential*.

EXERCISE 2.7. Show that grazing collisions are possible only on dispersing walls (where $\mathcal{K} > 0$). Hint: If the wall is flat, then before a grazing collision the particle should have passed an endpoint of the wall, which is prohibited by our rules.

EXERCISE 2.8. Let the moving particle collide with the regular part of the boundary at time t (that is $q(t) \in \tilde{\Gamma}$). Show that it will move inside $\mathcal{D}$ without collisions during some time interval $(t, t + \varepsilon)$.

We now summarize our observations:

PROPOSITION 2.9. *The trajectory of the particle $(q(t), v(t))$ starting at $q(0) \in \operatorname{int}\mathcal{D}$ 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 $\mathbb{R}$.*

2.4. Accumulation of collision times

The second exception mentioned in Proposition 2.9 may occur only under very special circumstances. In fact, it never happens in any chaotic billiards discussed in this book. Here we describe the conditions under which (b) is possible.

Suppose the event (b) occurs; i.e. the collision times $\{t_n\}$ have an accumulation point $t_n \to t_\infty \in \mathbb{R}$. Then, obviously, $q(t_n) \to q_\infty \in \Gamma$, and we distinguish two cases:

(b1) $q_\infty \in \Gamma_*$ (a corner point);
(b2) $q_\infty \in \tilde{\Gamma}$ (a regular point).

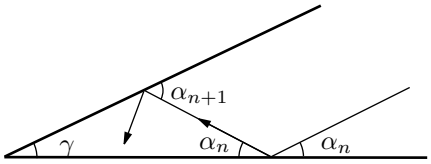

FIGURE 2.6. Proof of Lemma 2.10.

LEMMA 2.10. *Suppose the billiard particle enters the neighborhood of a corner point of the billiard table with a positive interior angle, $\gamma > 0$, and collides with both sides of the corner. Then it must leave that neighborhood after at most $[\pi/\gamma] + 1$ collisions, so that (b1) cannot occur.*

The case where the particle collides with only one side of the corner is equivalent to (b2) and will be discussed later.

PROOF. If the sides of the corner are flat, the angle made by the links of the billiard trajectory with the sides grows as $\alpha_{n+1} = \alpha_n + \gamma$; see Fig. 2.6. Thus, after $[\pi/\gamma]$ collisions it becomes $> \pi$, which proves the lemma. If the sides are not flat, then the result follows by a simple approximation argument, which we leave to the reader. □

Lemma 2.10 shows that (b1) may occur only in a cusp. Now, it is not hard to see that any cusp has at least one dispersing side (check this!).

LEMMA 2.11. *If the billiard particle enters a cusp with two dispersing sides or one dispersing and one flat side, then it must exit the cusp after a finite number of collisions, so that (b1) cannot occur.*

PROOF. Let d denote the distance from the corner point of the cusp to the line along which the particle moves. After every collision with a dispersing side that distance increases (see Fig. 2.7), while at collisions with a flat side it does not change. The rest of the proof requires some simple reasoning, which is left as an exercise. □

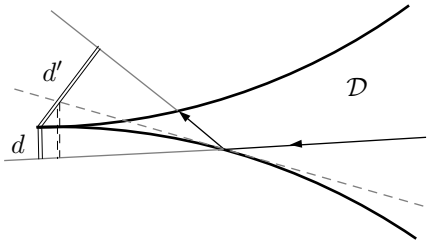

FIGURE 2.7. Proof of Lemma 2.11: $d' > d$.

EXERCISE 2.12. Show that the number of collisions in a cusp is not uniformly bounded; i.e. for any $N \geq 1$ there is a billiard trajectory than experiences more that N collisions in a vicinity of the cusp before leaving it (in contrast to the uniform bound for corners with positive angles provided by Lemma 2.10).

Thus, (b1) may be possible only in a cusp with one dispersing and one focusing side. But is it really possible? To our best knowledge, there are no published examples of this sort. We leave this question to the reader as a (possibly, very difficult) exercise. In any case, billiard tables with such cusps are very special, and we do not want to consider them in this book. To exclude them, we are making our last assumption:

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

We now turn to the type (b2) accumulations, i.e. $q_\infty \in \tilde{\Gamma}$. Clearly, (b2) is impossible if q_∞ lies on a flat or dispersing boundary component (check this!).

THEOREM 2.13 (B. Halpern [**Hal77**]). *The type (b2) accumulation of collisions is impossible on any focusing wall with a bounded third derivative and nowhere vanishing curvature.*

Halpern's argument is elementary but rather lengthy, so we omit it and refer the reader to his original paper [**Hal77**]. We note that Halpern found an example of a C^2 (not C^3) focusing billiard wall where the event (b2) does occur.

Due to Halpern's theorem, (b2) never occurs for the billiard tables defined in Section 2.1. Together with Assumption A4 made above, accumulations of collision points are now completely ruled out.

2.5. 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 velocity 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 Ω as a "doughnut" whose cross-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 (2.8), which amounts to "gluing" Ω along its boundary. This changes the topology of Ω, of course, but its topological properties will not be essential.

We denote by π_q and π_v the natural projections of Ω onto $\mathcal{D}$ and S^1, respectively, so that $\pi_q(q, v) = q$ and $\pi_v(q, v) = v$.

Let $\tilde{\Omega} \subset \Omega$ denote the set of states (q, v) on which the dynamics of the moving particle is defined at all times $-\infty < t < \infty$ (this set will be accurately described later). Thus we obtain a one-parameter group of transformations (flow)

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

with continuous time $t \in \mathbb{R}$. That is, $\Phi^0 = \mathrm{Id}$ and $\Phi^{t+s} = \Phi^t \circ \Phi^s$ for all $t, s \in \mathbb{R}$.

Every trajectory of the flow $\{\Phi^t x\}$, $x \in \tilde{\Omega}$, is a continuous curve in Ω (the continuity at collisions is ensured by our identification of v^+ with v^-; see above). It is also customary to call its projection $\pi_q(\Phi^t x)$ onto the table $\mathcal{D}$ a *billiard trajectory*. The latter is a directed polygonal line whose vertices are points of collision. Its segments between successive collisions are called *links*.

If the billiard table is a bounded domain $\mathcal{D} \subset \mathbb{R}^2$, then the free path between collisions cannot exceed the diameter of $\mathcal{D}$; hence every trajectory of the flow experiences infinitely many collisions. This may not be so if $\mathcal{D} \subsetneq \mathrm{Tor}^2$, as in the example shown in Fig. 2.4.

LEMMA 2.14. *If $\mathcal{D}$ is a billiard table on* Tor^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 (i.e., a straight line) on Tor^2 is either periodic or dense (the latter means that its closure is Tor^2); see Section 1.2. The rest of the proof is left as an exercise. □

Let $\tilde{\Omega} = \tilde{\Omega}_{\mathrm{c}} \cup \tilde{\Omega}_{\mathrm{f}}$, where $\tilde{\Omega}_{\mathrm{c}}$ contains all trajectories *with* collisions and $\tilde{\Omega}_{\mathrm{f}}$ is the union of all collision-free trajectories (which may exist only if $\mathcal{D} \subsetneq \mathrm{Tor}^2$). Note that both $\tilde{\Omega}_{\mathrm{c}}$ and $\tilde{\Omega}_{\mathrm{f}}$ are invariant under Φ^t. The rest of this section is devoted to billiard tables $\mathcal{D} \subsetneq \mathrm{Tor}^2$, for which $\tilde{\Omega}_{\mathrm{f}}$ may exist.

EXERCISE 2.15. Suppose $\tilde{\Omega}_{\mathrm{f}} \neq \emptyset$. Show that there are only finitely many velocity vectors $v \in S^1$ that correspond to points $x \in \tilde{\Omega}_{\mathrm{f}}$.

DEFINITION 2.16. We say that a billiard table $\mathcal{D} \subsetneq \mathrm{Tor}^2$ has *finite horizon* (or *bounded horizon*) if $\tilde{\Omega}_{\mathrm{f}} = \emptyset$. Otherwise the horizon is said to be *infinite* (or *unbounded*).

EXERCISE 2.17. Suppose $\tilde{\Omega}_{\mathrm{f}} = \emptyset$; i.e. the horizon is finite. Show that the free path between collisions is bounded above by some constant $\tau_{\max}(\mathcal{D}) < \infty$. Hints: Suppose the free path between collisions is unbounded; then there is a sequence of

line segments $L_n \subset \mathcal{D}$ of increasing length, $|L_n| \to \infty$. By the compactness of Tor^2, a subsequence L_{n_k} will converge to a line $L \subset \mathcal{D}$, which is obviously a periodic line on Tor^2. Since the horizon is finite, the line L must intersect the boundary $\partial\mathcal{D}$. Now, if at all those intersections the boundary $\partial\mathcal{D}$ lies on one side of L, then a nearby line parallel to L will have no intersections with $\partial\mathcal{D}$ at all, which contradicts the finite horizon assumption. Thus $\partial\mathcal{D}$ lies on both sides of L at their intersections. Now argue that the lengths of L_{n_k} must be bounded.

EXERCISE 2.18. Suppose $\mathcal{D}$ has no cusps; i.e. the interior angles at corner points are all positive. Show that the number of collisions per unit time is bounded by some $M_{\max} < \infty$. Hint: Use Lemma 2.10.

2.6. Coordinate representation of the flow

Here we describe the flow Φ^t in coordinates (x, y, ω) on Ω, where $q = (x, y) \in \mathcal{D}$. Here x and y are the usual Cartesian coordinates on the plane and $\omega \in [0, 2\pi)$ denotes the counterclockwise angle between the positive x axis and the velocity vector v. For every $t \in \mathbb{R}$ the map Φ^t acts on Ω, and we consider an arbitrary point and its image:

$$(x^-, y^-, \omega^-) \xrightarrow{\Phi^t} (x^+, y^+, \omega^+).$$

Our aim is to compute the derivative of the map Φ^t.

If no collisions occur, then by (2.7)

$$x^+ = x^- + t\cos\omega, \quad y^+ = y^- + t\sin\omega, \quad \omega^+ = \omega^-. \tag{2.9}$$

Suppose now there is exactly one (regular) collision at some Γ_i during the interval $(0, t)$. Then a direct relation between (x^-, y^-, ω^-) and (x^+, y^+, ω^+) will be rather complicated. To simplify matters, we will introduce some intermediate variables.

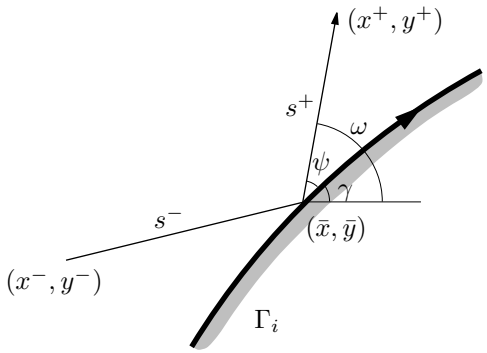

FIGURE 2.8. Action of the flow in coordinates.

Let $(\bar{x}, \bar{y}) \in \Gamma_i$ denote the collision point, $\mathbf{T}$ the tangent vector to Γ_i at that point, and γ the angle between $\mathbf{T}$ and the positive x axis. Let s^- denote the time of collision, $s^+ = t - s^-$, and ψ denote the angle between v^+ and $\mathbf{T}$; see Fig. 2.8.

Then

$$\begin{aligned} x^- &= \bar{x} - s^- \cos\omega^-, & x^+ &= \bar{x} + s^+ \cos\omega^+, \\ y^- &= \bar{y} - s^- \sin\omega^-, & y^+ &= \bar{y} + s^+ \sin\omega^+, \\ \omega^- &= \gamma - \psi, & \omega^+ &= \gamma + \psi. \end{aligned} \tag{2.10}$$

We are going to differentiate these equations. Let r denote the arc length parameter on Γ_i, then

$$\begin{aligned} d\bar{x} &= \cos\gamma\, dr, \\ d\bar{y} &= \sin\gamma\, dr, \\ d\gamma &= -\mathcal{K}\, dr \end{aligned}$$

(the sign of $d\gamma$ is completely determined by our orientation of Γ_i, as defined in Section 2.1).

Now differentiating (2.10) gives

$$\begin{aligned} dx^+ &= \cos\gamma\, dr + \cos\omega^+\, ds^+ - s^+ \sin\omega^+\, d\omega^+ \\ dy^+ &= \sin\gamma\, dr + \sin\omega^+\, ds^+ + s^+ \cos\omega^+\, d\omega^+ \\ d\omega^+ &= -\mathcal{K}\, dr + d\psi \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} dx^- &= \cos\gamma\, dr - \cos\omega^-\, ds^- + s^- \sin\omega^-\, d\omega^- \\ dy^- &= \sin\gamma\, dr - \sin\omega^-\, ds^- - s^- \cos\omega^-\, d\omega^- \\ d\omega^- &= -\mathcal{K}\, dr - d\psi. \end{aligned} \tag{2.12}$$

Exercise 2.19. Prove that

$$dx^+ \wedge dy^+ \wedge d\omega^+ = \sin\psi\ dr \wedge ds^+ \wedge d\psi \tag{2.13}$$

and

$$dx^- \wedge dy^- \wedge d\omega^- = -\sin\psi\ dr \wedge ds^- \wedge d\psi,$$

where $\wedge$ denotes the exterior multiplication of differential forms [**Ar89**, Chapter 7]. (It is convenient to use the skew-commutativity $du \wedge dv = -dv \wedge du$, and in particular $du \wedge du = 0$.) The reader unfamiliar with exterior forms may, instead, directly compute the Jacobians of the maps $(r, s, \psi) \mapsto (x^\pm, y^\pm, \omega^\pm)$ and show that they are equal to $\pm\sin\psi$, respectively.

Note that $s^- + s^+ = t = \text{const}$; hence $ds^+ + ds^- = 0$, and we obtain

$$dx^+ \wedge dy^+ \wedge d\omega^+ = dx^- \wedge dy^- \wedge d\omega^-. \tag{2.14}$$

Exercise 2.20. Use (2.9) to show that if there are no collisions between the points (x^-, y^-, ω^-) and (x^+, y^+, ω^+), then (2.14) holds as well.

Now an induction on the number of collisions implies the following:

Theorem 2.21. *The flow Φ^t preserves the volume form $dx \wedge dy \wedge d\omega$; thus it preserves the Lebesgue measure $dx\, dy\, d\omega$ on Ω.*

Remark 2.22. The Lebesgue measure on Ω is actually the Liouville measure corresponding to the Hamiltonian system (2.7)–(2.8). This fact is often given as a "proof" of the invariance of the Lebesgue measure; however, we avoided this argument because it is hard to check that the collision rule (2.8) preserves the Hamiltonian character of motion.

DEFINITION 2.23. The normalized Lebesgue measure on Ω

$$d\mu_\Omega = (2\pi|\mathcal{D}|)^{-1}\, dx\, dy\, d\omega \tag{2.15}$$

(here $|\mathcal{D}|$ denotes the area of the domain $\mathcal{D}$) is the canonical probability measure preserved by the billiard flow Φ^t.

2.7. Smoothness of the flow

LEMMA 2.24. *The flow Φ^t is $C^{\ell-1}$ smooth at points that experience only regular collisions.*

PROOF. Equations (2.11)–(2.12) show that the derivative of the flow Φ^t is expressed through the curvature $\mathcal{K}$ of the wall Γ_i, which in turn corresponds to the second derivative of the C^ℓ function $f_i(r)$. □

FIGURE 2.9. A grazing collision; here the flow is continuous (but not differentiable).

Now suppose there is a grazing collision of a billiard trajectory between the points (x^-, y^-, ω^-) and (x^+, y^+, ω^+) at a dispersing wall. Then it is not hard to see that Φ^t is not differentiable at (x^-, y^-, ω^-) (even though it is continuous; see Fig. 2.9). We leave this as an exercise.

The set $\Omega \setminus \tilde{\Omega}$ (see Section 2.5 for the definition of $\tilde{\Omega}$) corresponds to states (q, v) on which the dynamics is defined only during a limited time interval until the trajectory hits a corner point $q' \in \Gamma_*$ (in the past or in the future). The states $x \in \Omega$ whose very first contact with Γ (in the past or future) occurs at a corner point make a two-dimensional surface in Ω; see Fig. 2.10.

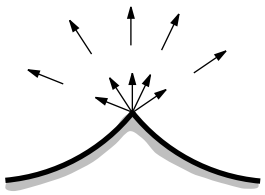

FIGURE 2.10. A "fan" emanating from a corner point of $\mathcal{D}$. These are phase points whose first contact with $\partial\mathcal{D}$ in the past occurs at the corner point.

Since the flow is $C^{\ell-1}$ at regular collisions, the states whose trajectories hit corner points in the distant past or distant future make two-dimensional $C^{\ell-1}$ submanifolds in Ω. Thus we obtain

THEOREM 2.25. *$\Omega \setminus \tilde{\Omega}$ is a countable union of two-dimensional submanifolds in Ω.*

COROLLARY 2.26. *$\tilde{\Omega}$ is a dense G_δ set of full Lebesgue measure in Ω.*

Thus, in terms of the invariant measure μ_Ω, the flow Φ^t is defined almost everywhere.

2.8. Continuous extension of the flow

Here we extend the flow Φ^t to the entire space Ω by continuity.

EXERCISE 2.27. Show that the flow Φ^t is continuous on $\tilde{\Omega}$ (assuming the identification of (q, v^-) and (q, v^+) as above). Warning: Grazing collisions at dispersing walls must be treated carefully; see Fig. 2.9.

Since $\tilde{\Omega}$ is a dense subset of Ω, we can extend the flow Φ^t to the entire space Ω by continuity, but this extension may be multivalued at some points; see an example in Fig. 2.11. Different extensions of Φ^t at the same point $x \in \Omega$ are called *branches* of the (extended) flow. For each $t \in \mathbb{R}$ there are finitely many branches of the extended flow Φ^t at any point $x \in \Omega$. The number of branches may grow indefinitely as $t \to \pm\infty$.

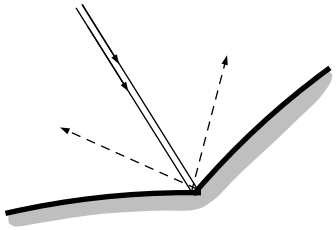

FIGURE 2.11. A multivalued continuous extension of the flow.

EXERCISE 2.28. Show that a trajectory hitting a corner point can be extended, by continuity, to at most two different branches. Under what conditions do these extensions coincide (i.e., when is the flow actually continuous) at a trajectory hitting a corner point? Answer: The branches coincide if and only if the interior angle at the corner point either equals zero or divides π (that is, $\gamma = 0$ or $\gamma = \pi/n$ for some $n \in \mathbb{N}$). Hints: For $\gamma > 0$, see the proof of Lemma 2.10. For $\gamma = 0$, see the proof of Lemma 2.11. In both cases a careful geometric analysis is required to complete the proof.

EXERCISE 2.29. Determine under what conditions the billiard flow Φ^t has a globally continuous extension to the entire space Ω. Answer: This occurs if and only if every interior angle either equals zero or divides π (as an example, recall the billiard in a square in Section 1.2).

2.9. Collision map

It is common in the study of dynamical systems to reduce a flow to a map by constructing a cross-section. Given a flow $\Phi^t\colon \Omega \to \Omega$ on a manifold Ω, 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 $F\colon M \to M$ and a return time function $L(x) = \min\{s > 0\colon \Phi^s(x) \in M\}$ on M, so that $F(x) = \Phi^{L(x)}(x)$.

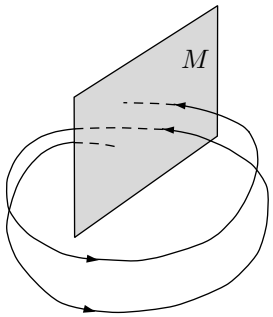

FIGURE 2.12. Cross-section of a flow.

Conversely, given a measurable space M, a measurable map $F\colon M \to M$, and a positive function $L\colon M \to \mathbb{R}_+$, one can construct a space

$$\Omega = \{(x, s)\colon x \in M,\ 0 \le s \le L(x)\}$$

and a flow $\Phi^t\colon \Omega \to \Omega$ defined by $\Phi^t(x, s) = (x, s + t)$ with the identification of points $(x, L(x))$ and $(F(x), 0)$. The flow Φ^t is measurable on Ω. If the map F preserves a probability measure μ on M and

$$\bar{L} = \int_M L(x)\, d\mu(x) < \infty,$$

then the flow preserves the probability measure μ_1 on Ω defined by

$$d\mu_1 = \bar{L}^{-1}\, d\mu \times ds \tag{2.16}$$

(note that this is, locally, a product measure whose invariance under Φ^t is a consequence of the Fubini theorem). The map F is often called *base transformation*, $L(x)$ *ceiling function*, and Φ^t *suspension flow* or *Kakutani flow*. We refer the reader to [**BrS02**, pp. 21–22], [**DS00**, p. 26–28], [**CFS82**, pp. 292–295], and [**Pet83**, p. 11] for more comprehensive discussions of these constructions.

For a billiard system, a hypersurface in Ω is usually constructed on the boundary of the billiard table, i.e. on the set $\Gamma \times S^1$. There is a delicate part, however, in this construction, since we have identified the precollisional and postcollisional velocity vectors v^- and v^+ related by (2.8), thus effectively transforming S^1 into a half-circle. To take this identification into account, it is customary to describe the cross-section as the set of all postcollisional velocity vectors:

$$\mathcal{M} = \cup_i \mathcal{M}_i, \quad \mathcal{M}_i = \{x = (q, v) \in \Omega\colon\ q \in \Gamma_i,\ \langle v, n\rangle \ge 0\}, \tag{2.17}$$

where n denotes the unit normal vector to Γ_i pointing inside $\mathcal{D}$. The set $\mathcal{M}$ is a two-dimensional submanifold in Ω called the *collision space*.

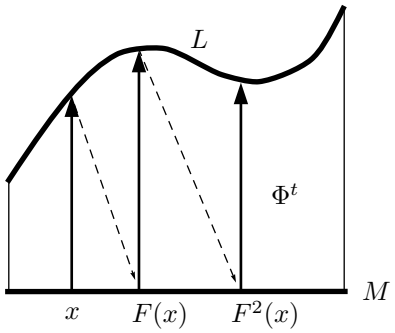

FIGURE 2.13. A suspension flow.

EXERCISE 2.30. Let $x = (q, v) \in \mathcal{M}_i$. Under what conditions is the trajectory $\Phi^t x$ defined during some small interval $0 < t < \varepsilon$? Answer: q must be a regular point ($q \in \tilde{\Gamma}$), and we must have either $\langle v, n\rangle > 0$ (a regular collision) or $\langle v, n\rangle = 0$, but in the latter case Γ_i must be a dispersing wall.

If the trajectory $\Phi^t x$ for $x \in \mathcal{M}$ is defined during some interval of time $(0, \varepsilon)$, then it must intersect the surface $\Gamma \times S^1$ at a future time $\tau(x) > 0$ according to Lemma 2.14, and we call $\tau(x)$ the *return time*. Since the speed is set to one, $\tau(x)$ will also be equal to the distance the billiard trajectory originating at x covers before the next collision.

Any trajectory of the flow $\Phi^t \colon \tilde{\Omega}_c \to \tilde{\Omega}_c$ (see Section 2.5 for the definition of $\tilde{\Omega}_c$) crosses the surface $\mathcal{M}$ infinitely many times. Let $\tilde{\mathcal{M}} = \mathcal{M} \cap \tilde{\Omega}$. This defines the *return map*

$$\mathcal{F} \colon \tilde{\mathcal{M}} \to \tilde{\mathcal{M}} \qquad \text{by} \qquad \mathcal{F}(x) = \Phi^{\tau(x)+0} x. \tag{2.18}$$

We will extend $\mathcal{F}$ to $\mathcal{M} \setminus \tilde{\mathcal{M}}$ later. The map $\mathcal{F}$ is often called the *billiard map* or *collision map*. (Accordingly, $\mathcal{M}$ is sometimes called the *phase space of the billiard map* $\mathcal{F}$.)

2.10. Coordinates for the map and its singularities

We fix an arclength parameter r on each Γ_i, so that r takes values in an interval $[a_i, b_i]$. Of course, $b_i - a_i = |\Gamma_i|$. We assume that the intervals (a_i, b_i) are disjoint in $\mathbb{R}$. On smooth closed curves Γ_i, we identify $a_i = b_i$, thus making r a cyclic parameter. For each point $x \in \mathcal{M}$, let $\varphi \in [-\pi/2, \pi/2]$ denote the angle between v and n oriented as shown in Fig. 2.14.

Then r and φ make coordinates on $\mathcal{M}$. For each smooth closed curve Γ_i, the manifold $\mathcal{M}_i = \Gamma_i \times [-\pi/2, \pi/2]$ is a cylinder (since r is a cyclic coordinate), while for every other boundary component Γ_i, the manifold $\mathcal{M}_i = [a_i, b_i] \times [-\pi/2, \pi/2]$ is a rectangle; see Fig. 2.15.

We denote

$$\mathcal{S}_0 \colon = \partial\mathcal{M} = \{|\varphi| = \pi/2\} \cup \Big(\cup_i \big(\{r = a_i\} \cup \{r = b_i\} \big) \Big), \tag{2.19}$$

where the set $\{r = a_i\} \cup \{r = b_i\}$ is included only for Γ_i's which are *not* smooth closed curves (i.e., where $\{a_i\} \cup \{b_i\}$ is a *true* boundary of the interval $[a_i, b_i]$).

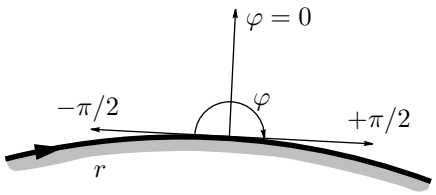

FIGURE 2.14. Orientation of r and φ.

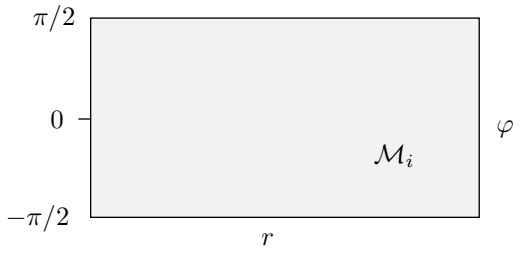

FIGURE 2.15. A component of the collision space $\mathcal{M}$.

For every point $x \in \operatorname{int}\mathcal{M}$ 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. $\mathcal{F}(x) \in \operatorname{int}\mathcal{M}$;
(b) a grazing collision at a dispersing wall, i.e. $\mathcal{F}(x) \in \mathcal{S}_0$;
(c) the trajectory hits a corner point and dies.

In the last case $\mathcal{F}(x)$ is not defined. Note that $\mathcal{S}_1 \setminus \mathcal{S}_0$ is the set of points where the event (b) or (c) occurs.

EXERCISE 2.31. Show that the set $\mathcal{M} \setminus \mathcal{S}_1$ is open and at every point of this set the map $\mathcal{F}$ is continuous and in fact is a local homeomorphism. Verify that at every point $x \in \mathcal{S}_1 \setminus \mathcal{S}_0$ the map $\mathcal{F}$ is discontinuous.

EXERCISE 2.32. Check that the above considerations can be applied to the inverse map $\mathcal{F}^{-1}$ and the set

$$\mathcal{S}_{-1} = \mathcal{S}_0 \cup \{x \in \operatorname{int}\mathcal{M} \colon \mathcal{F}^{-1}(x) \notin \operatorname{int}\mathcal{M}\}. \tag{2.20}$$

Conclude that $\mathcal{F} \colon \mathcal{M} \setminus \mathcal{S}_1 \to \mathcal{M} \setminus \mathcal{S}_{-1}$ is a homeomorphism.

2.11. Derivative of the map

Here we differentiate the collision map $\mathcal{F}$ at a point $x = (r, \varphi) \in \operatorname{int}\mathcal{M}$ such that $\mathcal{F}(x) = (r_1, \varphi_1) \in \operatorname{int}\mathcal{M}$. Denote by $(\bar{x}, \bar{y}) \in \partial\mathcal{D}$ and $(\bar{x}_1, \bar{y}_1) \in \partial\mathcal{D}$ the coordinates of the boundary points corresponding to r and r_1, respectively, and by ω the angle made by the billiards trajectory between these points and the positive x axis in $\mathbb{R}^2$; see Fig. 2.16. Then

$$\bar{x}_1 - \bar{x} = \tau\cos\omega \qquad \text{and} \qquad \bar{y}_1 - \bar{y} = \tau\sin\omega \tag{2.21}$$

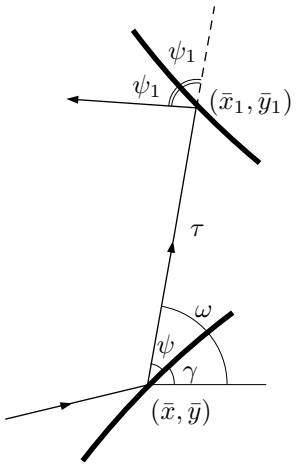

FIGURE 2.16. Action of the map in coordinates.

where $\tau = \tau(x)$. We use the symbols γ and ψ as they were introduced in Section 2.6; see also Fig. 2.16. Note that

$$\psi = \pi/2 - \varphi$$

and recall that

$$\begin{aligned} d\bar{x} &= \cos\gamma\, dr, \\ d\bar{y} &= \sin\gamma\, dr, \\ d\gamma &= -\mathcal{K}\, dr. \end{aligned}$$

Similar notations γ_1 and ψ_1 are used at the point r_1. Note that

$$\omega = \gamma + \psi = \gamma_1 - \psi_1. \tag{2.22}$$

Differentiating the last equation gives

$$d\omega = -\mathcal{K}\, dr + d\psi = -\mathcal{K}_1\, dr_1 - d\psi_1. \tag{2.23}$$

Differentiating (2.21) yields

$$\begin{aligned} \cos\gamma_1\, dr_1 - \cos\gamma\, dr &= \cos\omega\, d\tau - \tau\sin\omega\, d\omega \\ \sin\gamma_1\, dr_1 - \sin\gamma\, dr &= \sin\omega\, d\tau + \tau\cos\omega\, d\omega. \end{aligned}$$

Eliminating $d\tau$ and using (2.22) give

$$\sin\psi_1\, dr_1 + \sin\psi\, dr = \tau\, d\omega. \tag{2.24}$$

Solving the equations (2.23) and (2.24) for dr_1 and $d\psi_1$ and replacing ψ and ψ_1 by $\pi/2 - \varphi$ and $\pi/2 - \varphi_1$, respectively, give

$$\begin{aligned} -\cos\varphi_1\, dr_1 &= (\tau\mathcal{K} + \cos\varphi)\, dr + \tau\, d\varphi \\ -\cos\varphi_1\, d\varphi_1 &= (\tau\mathcal{K}\mathcal{K}_1 + \mathcal{K}\cos\varphi_1 + \mathcal{K}_1\cos\varphi)\, dr \\ &\quad + (\tau\mathcal{K}_1 + \cos\varphi_1)\, d\varphi. \end{aligned} \tag{2.25}$$

Thus we obtain the derivative $D\mathcal{F}$ at the point $x = (r, \varphi)$ as a 2×2 matrix

$$D_x\mathcal{F} = \frac{-1}{\cos\varphi_1} \begin{bmatrix} \tau\mathcal{K} + \cos\varphi & \tau \\ \tau\mathcal{K}\mathcal{K}_1 + \mathcal{K}\cos\varphi_1 + \mathcal{K}_1\cos\varphi & \tau\mathcal{K}_1 + \cos\varphi_1 \end{bmatrix}. \tag{2.26}$$

Recall that the map $\mathcal{F}\colon \mathcal{M} \setminus \mathcal{S}_1 \to \mathcal{M} \setminus \mathcal{S}_{-1}$ is a homeomorphism. Now we claim more:

THEOREM 2.33. *The map $\mathcal{F}\colon \mathcal{M} \setminus \mathcal{S}_1 \to \mathcal{M} \setminus \mathcal{S}_{-1}$ is a $C^{\ell-1}$ diffeomorphism.*

PROOF. The derivative $D\mathcal{F}$ is expressed through the curvature $\mathcal{K}$ and $\mathcal{K}_1$ of the boundary $\partial\mathcal{D}$, which corresponds to the second derivative of the C^ℓ functions $f_i\colon [a_i, b_i] \to \mathbb{R}^2$. □

The expression (2.26) shows that the derivatives of $\mathcal{F}$ are unbounded: they blow up as $\cos\varphi_1 \to 0$, i.e. when x_1 is near $\mathcal{S}_0$ and x is near $\mathcal{S}_1$.

We regard $\mathcal{S}_1$ as the *singularity set* for the map $\mathcal{F}$. Similarly, $\mathcal{S}_{-1}$ is the *singularity set* for the map $\mathcal{F}^{-1}$. These sets are unions of compact smooth curves (their exact degree of smoothness will be determined later); cf. Theorem 2.25.

EXERCISE 2.34. Suppose the horizon is bounded. Show that $\mathcal{S}_1$ and $\mathcal{S}_{-1}$ are *finite* unions of compact smooth curves. If the horizon is unbounded, then $\mathcal{S}_1$ and $\mathcal{S}_{-1}$ are *finite* or *countable* unions of compact smooth curves.

We define, inductively,

$$\mathcal{S}_{n+1} = \mathcal{S}_n \cup \mathcal{F}^{-1}(\mathcal{S}_n) \quad \text{and} \quad \mathcal{S}_{-(n+1)} = \mathcal{S}_{-n} \cup \mathcal{F}(\mathcal{S}_{-n}). \tag{2.27}$$

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. These sets are unions of compact smooth curves. Thus, on the set

$$\hat{\mathcal{M}} \colon = \mathcal{M} \setminus \cup_{n=-\infty}^{\infty} \mathcal{S}_n \tag{2.28}$$

all the iterations of $\mathcal{F}$ are defined and are $C^{\ell-1}$ differentiable.

As a result, the map $\mathcal{F}$ is well defined by (2.18) on a dense G_δ subset $\hat{\mathcal{M}} \subset \mathcal{M}$ of full Lebesgue measure. It can be extended by continuity to the entire $\mathcal{M}$, but this extension may be multivalued at some points.

2.12. Invariant measure of the map

By using the explicit formula for $D_x\mathcal{F}$ (see (2.26)), one easily obtains

$$\det D_x\mathcal{F} = \cos\varphi / \cos\varphi_1. \tag{2.29}$$

LEMMA 2.35. *The map $\mathcal{F}$ preserves the measure $\cos\varphi\, dr\, d\varphi$ on $\mathcal{M}$.*

PROOF. Using (2.29) and changing variables give

$$\iint\limits_{\mathcal{F}(A)} \cos\varphi_1\, dr_1\, d\varphi_1 = \iint\limits_{A} \cos\varphi\, dr\, d\varphi$$

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

Note that

$$\iint\limits_{\mathcal{M}} \cos\varphi\, dr\, d\varphi = \int_{-\pi/2}^{\pi/2} \cos\varphi\, d\varphi \int_\Gamma dr = 2\,|\Gamma|.$$

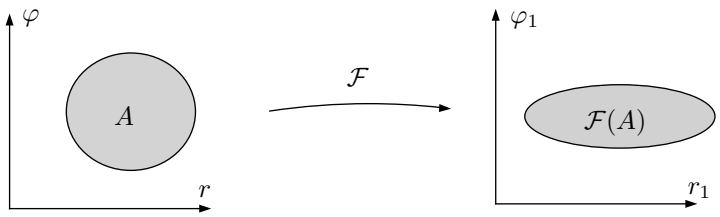

FIGURE 2.17. Proof of Lemma 2.35.

DEFINITION 2.36. The normalized measure on $\mathcal{M}$,

$$d\mu = (2\,|\Gamma|)^{-1}\,\cos\varphi\,dr\,d\varphi,$$

is the canonical probability measure preserved by the billiard map $\mathcal{F}$.

According to the general fact (2.16), the flow Φ^t must preserve the probability measure

$$\begin{aligned} d\mu_1 &= \bar\tau^{-1}\,d\mu\times ds \\ &= (2\,|\Gamma|)^{-1}\,\bar\tau^{-1}\cos\varphi\,dr\,d\varphi\,ds, \end{aligned} \tag{2.30}$$

where

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

is the *mean return time*. The latter is more often called the *mean free path* on the billiard table $\mathcal{D}$ (note that time=path, since the velocity equals one). Due to (2.13) we can rewrite (2.30) as

$$d\mu_1 = (2\,|\Gamma|)^{-1}\,\bar\tau^{-1} dx\,dy\,d\omega$$

(here we used the fact $\sin\psi = \cos\varphi$). Comparing this formula with (2.15) we conclude that $\mu_1 = \mu_\Omega$ and, therefore, the normalizing coefficients for these measures coincide:

$$2\pi|\mathcal{D}| = 2\,|\Gamma|\,\bar\tau \tag{2.31}$$

(by the way, this also proves that $\bar\tau < \infty$).

EXERCISE 2.37. Show that for μ-almost every point $x\in\mathcal{M}$ there exist limits

$$\bar\tau_x = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\tau(\mathcal{F}^i x) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\tau(\mathcal{F}^{-i} x),$$

where $\bar\tau_x$ is an $\mathcal{F}$-invariant function on $\mathcal{M}$ satisfying

$$\int_{\mathcal{M}}\bar\tau_x\,d\mu = \bar\tau.$$

Hint: Use the Birkhoff ergodic theorem.

EXERCISE 2.38. Show that for μ_Ω-almost every point $X\in\Omega$ there exist limits

$$\bar\tau_X = \lim_{T\to\infty}\frac{T}{n_X(0,T)} = \lim_{T\to\infty}\frac{T}{n_X(-T,0)}$$

where $n_X(a,b)$ denotes the number of collisions of the trajectory $\Phi^t(X)$ during the time interval $a < t < b$. Moreover, show that $\bar\tau_X = \bar\tau_x$ if the points $X \in \Omega$ and $x \in \mathcal{M}$ belong to the same trajectory.

2.13. Mean free path

Equation (2.31) gives us a simple and elegant formula for the mean free path on a billiard table $\mathcal{D}$:

$$\bar\tau = \frac{\pi|\mathcal{D}|}{|\Gamma|}. \tag{2.32}$$

Thus, the mean free path depends only on the area and perimeter of the domain $\mathcal{D}$, but not on its shape. This equation is well known in integral geometry and geometric probability, where it is proved by different methods [**Mat75, Sa76**].

EXERCISE 2.39. Derive (2.32) by a direct integration of the identity (2.13). Hint: Note that $\int ds^+ = \tau$.

EXERCISE 2.40. Compute the mean free path on a billiard table $\mathcal{D} = \mathrm{Tor}^2 \setminus B_{\mathbf{r}}$ where $B_{\mathbf{r}}$ is a ball (disk) of radius $\mathbf{r}$; see Fig. 2.4. Answer: $\bar\tau = (1 - \pi\mathbf{r}^2)/(2\mathbf{r})$. Note that if $\mathbf{r}$ is small, then $\bar\tau \approx 1/(2\mathbf{r})$. This can be understood in probabilistic terms: the billiard particle moving freely on the unit torus may randomly hit the fixed disk $B_{\mathbf{r}}$ with probability equal to its diameter $2\mathbf{r}$; thus the average distance between collisions is $\approx 1/(2\mathbf{r})$.

If the horizon is bounded, then $\tau(x) \le \tau_{\max} < \infty$; hence all the moments of the function $\tau(x)$ are finite. This may not be so when the horizon is unbounded, i.e. when $\sup \tau(x) = \infty$.

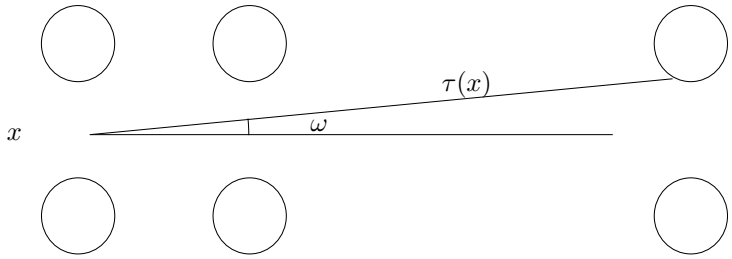

FIGURE 2.18. Hint to Exercise 2.41.

EXERCISE 2.41. Consider the billiard table of Exercise 2.40. Show that

$$\int_{\mathcal{M}} [\tau(x)]^2 \, d\mu(x) = \infty.$$

Hints: First, use (2.13) to replace the above integral with

$$(2\,|\Gamma|)^{-1} \int_{\mathcal{M}} [\tau(r,\varphi)]^2 \, \cos\varphi \, dr \, d\varphi = (2\,|\Gamma|)^{-1} \int_{\Omega} \tau(x,y,\omega) \, dx \, dy \, d\omega,$$

where $\tau(x,y,\omega)$ denotes the length of the link of the billiard trajectory passing through the point $(x,y,\omega) \in \Omega$ (the link between the first future collision and the last past collision). Now consider "almost horizontal trajectories" with $\omega \approx 0$

passing through points $(x, y) \in \mathcal{D}$ located above or below the disk $B_{\mathbf{r}}$. Observe that $\tau(x, y, \omega) \sim 1/\omega$; thus the above integral diverges.

The function $\tau(x)$ may have a positive lower bound, as in the last two exercises. However, if the billiard table has a focusing boundary component or a corner point with interior angle $\gamma < \pi$, then $\min_x \tau(x) = 0$. In that case the distribution of small values of $\tau(x)$ is important, and the reader is invited to verify the following facts.

EXERCISE 2.42. Show that if the billiard table has a focusing boundary component but no corner points, then $\mu\{x\colon \tau(x) < \varepsilon\} \le \text{const} \cdot \varepsilon^2$. If the billiard table has a corner point with interior angle $\gamma < \pi$ but no cusps, then $\mu\{x\colon \tau(x) < \varepsilon\} \le \text{const} \cdot \varepsilon$. If the table has a cusp, then $\mu\{x\colon \tau(x) < \varepsilon\} \le \text{const} \cdot \varepsilon^{1/2}$. The last case is quite difficult; the reader should not expect an easy solution.

EXERCISE 2.43. Determine the set of positive numbers $a > 0$ for which the function $F(x) = [\tau(x)]^{-a}$ is integrable over $\mathcal{M}$.

Answer: For the three cases described in the previous exercise, these intervals are $(0, 2)$, $(0, 1)$, and $(0, 1/2)$, respectively.

2.14. Involution

Billiard dynamics has an important *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. Hence, the involution $\mathcal{I}_\Omega\colon \Omega \to \Omega$ anticommutes with the flow Φ^t, and this fact can be written as

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

This means, plainly, that if we reverse the particle's velocity, it will retrace its past trajectory 'backwards' (this fact is known as *time reversibility* of the billiard dynamics). Note that the map $\mathcal{I}_\Omega$ also preserves the measure μ_Ω.

The collision map $\mathcal{F}$ also admits an involution, $\mathcal{I}$, defined by $(r, \varphi) \mapsto (r, -\varphi)$. It anticommutes with $\mathcal{F}$; i.e.

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

whenever the map $\mathcal{F}^k$ is defined. We also note that the map $\mathcal{I}\colon \mathcal{M} \to \mathcal{M}$ preserves the measure μ.

FIGURE 2.19. Involution of the map $\mathcal{F}$.

EXERCISE 2.44. Consider the vector function $\mathbf{L}\colon \mathcal{M} \to \mathbb{R}^2$ defined by

$$\mathbf{L}(r,\varphi) = (\bar{x}_1 - \bar{x}, \bar{y}_1 - \bar{y}) = (\tau\cos\omega, \tau\sin\omega)$$

as illustrated by Fig. 2.16. This is the first directed link of the billiard trajectory originating from (r,φ). Show that

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

Hint: Note that $\mathbf{L}(x) = -\mathbf{L}(\mathcal{I}\circ\mathcal{F}(x))$ and then use the invariance of the measure μ under both $\mathcal{F}$ and $\mathcal{I}$.

Bibliography

[Ab59] L. M. Abramov, *On the entropy of a flow* (in Russian), Dokl. Akad. Nauk SSSR **128** (1959), 873–875.

[Al69] V. M. Alekseev, *Quasi random dynamical systems*, Mat. USSR Sbornik **7** (1969), 1–43.

[An67] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds with negative curvature*, Proc. Steklov Inst. Math. **90** (1967).

[Ar89] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer, New York, 1989.

[AS67] D. V. Anosov & Ya. G. Sinai, *Some smooth ergodic systems*, Russ. Math. Surv. **22** (1967), 103–167.

[AW67] R. Adler & B. Weiss, *Entropy, a complete metric invariant for automorphisms of the torus*, Proc. Nat. Acad. Sci. USA **57** (1967), 1573–1576.

[Ba00] V. Baladi, *Positive transfer operators and decay of correlations*, World Scientific, River Edge, NY, 2000.

[BCST02] P. Balint, N. Chernov, D. Szasz, and I. P. Toth, *Multi-dimensional semi-dispersing billiards: singularities and the fundamental theorem*, Ann. H. Poincaré **3** (2002), 451–482.

[BCST03] P. Balint, N. Chernov, D. Szasz, and I. P. Toth, *Geometry of multidimensional dispersing billiards*, Astérisque **286** (2003), 119–150.

[BD87] R. Burton & M. Denker, *On the central limit theorem for dynamical systems*, Trans. AMS **302** (1987), 715–726.

[Be01] A. Berger, *Chaos and chance*, Walter de Gruyter, Berlin, 2001.

[BG06] P. Balint & S. Gouezel, *Limit theorems in the stadium billiard*, preprint.

[Bi27] G. D. Birkhoff, *Dynamical systems*, A.M.S. Colloquium Publications, New York (1927).

[Bl92] P. M. Bleher, *Statistical properties of two-dimensional periodic Lorentz gas with infinite horizon*, J. Stat. Phys. **66** (1992), 315–373.

[BLPS92] L. A. Bunimovich, C. Liverani, A. Pellegrinotti & Y. Suhov, *Ergodic systems of n balls in a billiard table*, Comm. Math. Phys. **146** (1992), 357–396.

[BL02] X. Bressaud & C. Liverani, *Anosov diffeomorphism and coupling*, Ergod. Th. Dynam. Syst. **22** (2002), 129–152.

[Bou60] N. Bourbaki, *Éléments de mathématique*. XXV. Part 1. Livre VI: Intégration. Chap. 6: Intégration vectorielle. Actualites scientifiques et industrielles, **1281**, Paris: Hermann (1960).

[Bow70] R. Bowen, *Markov partitions for Axiom A diffeomorphisms*, Amer. J. Math. **92** (1970), 725–747.

[Bow75] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes Math. **470**, Springer, Berlin, 1975.

[BP01] L. Barreira & Ya. Pesin, *Lyapunov exponents and smooth ergodic theory*, Univ. Lect. Series, **23**, AMS, Providence, RI, 2001.

[BrS02] M. Brin & G. Stuck, *Introduction to Dynamical Systems*, Cambridge U. Press, 2002.

[BS73] L. A. Bunimovich & Ya. G. Sinai, *On a fundamental theorem in the theory of dispersing billiards*, Math. USSR Sbornik **19** (1973), 407–423.

[BS80] L. A. Bunimovich & Ya. G. Sinai, *Markov partitions for dispersed billiards*, Comm. Math. Phys. **78** (1980), 247–280.

[BS81] L. A. Bunimovich & Ya. G. Sinai, *Statistical properties of Lorentz gas with periodic configuration of scatterers*, Comm. Math. Phys. **78** (1981), 479–497.

[BS96] L. A. Bunimovich & H. Spohn, *Viscosity for a periodic two disk fluid: an existence proof*, Comm. Math. Phys. **176** (1996), 661–680.

[BR97] L. A. Bunimovich & J. Rehácek, *Nowhere dispersing* 3*D billiards with non-vanishing Lyapunov exponents*, Comm. Math. Phys. **189** (1997), 729–757.

[BR98a] L. A. Bunimovich & J. Rehácek, *On the ergodicity of many-dimensional focusing billiards. Classical and quantum chaos*, Ann. H. Poincaré **68** (1998), 421–448.

[BR98b] L. A. Bunimovich & J. Rehácek, *How high-dimensional stadia look like*, Comm. Math. Phys. **197** (1998), 277–301.

[BSC90] L. A. Bunimovich, Ya. G. Sinai & N. I. Chernov, *Markov partitions for two-dimensional billiards*, Russ. Math. Surv. **45** (1990), 105–152.

[BSC91] L. A. Bunimovich, Ya. G. Sinai & N. I. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Russ. Math. Surv. **46** (1991), 47–106.

[Bu74a] L. A. Bunimovich, *On billiards close to dispersing*, Math. USSR. Sb. **23** (1974), 45–67.

[Bu74b] L. A. Bunimovich, *The ergodic properties of certain billiards*, Funk. Anal. Prilozh. **8** (1974), 73–74.

[Bu79] L. A. Bunimovich, *On ergodic properties of nowhere dispersing billiards*, Comm. Math. Phys. **65** (1979), 295–312.

[Bu90] L. A. Bunimovich, *A theorem on ergodicity of two-dimensional hyperbolic billiards*, Comm. Math. Phys. **130** (1990), 599–621.

[Bu92] L. A. Bunimovich, *On absolutely focusing mirrors*, In *Ergodic Theory and related topics, III (Güstrow, 1990)*. Edited by U. Krengel et al., Lecture Notes in Math. **1514**, Springer, Berlin (1992), 62–82.

[Bu00] L. A. Bunimovich, *Hyperbolicity and astigmatism*, J. Stat. Phys. **101** (2000), 373–384.

[BY00] M. Benedicks & L.-S. Young, *Markov extensions and decay of correlations for certain Hénon maps*, Astérisque **261** (2000), 13–56.

[CCG] G. Casati, G. Comparin & I. Guarneri, *Decay of correlations in certain hyperbolic systems*, Phys. Rev. A **26** (1982), 717–719.

[CD06] N. Chernov & D. Dolgopyat, *Brownian Brownian motion – I*, to appear in Memoir. AMS.

[CFS82] I. P. Cornfeld, S. V. Fomin & Ya. G. Sinai, *Ergodic theory*, Fundamental Principles of Mathematical Sciences, **245**, Springer-Verlag, New York, 1982.

[Ch93] N. I. Chernov, *On local ergodicity in hyperbolic systems with singularities*, Funct. Anal. Appl. **27** (1993), 51–54.

[Ch95] N. Chernov, *Limit theorems and Markov approximations for chaotic dynamical systems*, Prob. Th. Rel. Fields **101** (1995), 321–362.

[CH96] N. Chernov & C. Haskell, *Nonuniformly hyperbolic K-systems are Bernoulli*, Ergod. Th. Dynam. Syst. **16** (1996), 19–44.

[Ch97] N. Chernov, *Entropy, Lyapunov exponents, and mean free path for billiards*, J. Stat. Phys. **88** (1997), 1–29.

[Ch99] N. Chernov, *Decay of correlations and dispersing billiards*, J. Stat. Phys. **94** (1999), 513–556.

[Ch01] N. Chernov, *Sinai billiards under small external forces*, Ann. H. Poincaré **2** (2001), 197–236.

[Ch02] N. Chernov, *Invariant measures for hyperbolic dynamical systems*, In: Handbook of Dynamical Systems, **1A**, 321–407, North-Holland, Amsterdam, 2002.

[Ch06a] N. Chernov, *Regularity of local manifolds in dispersing billiards*, Math. Phys. Electr. J. **12** (2006), No. 1, 54 pp.

[Ch06b] N. Chernov, *Advanced statistical properties of dispersing billiards*, to appear in J. Stat. Phys. (2006).

[CM92] N. I. Chernov & R. Markarian, *Entropy of non-uniformly hyperbolic plane billiards*, Bol. Soc. Bras. Mat. **23** (1992) 121–135.

[CM03] N. Chernov & R. Markarian, *Introduction to the ergodic theory of chaotic billiards*, 2nd Ed., IMPA, Rio de Janeiro, Brasil, 2003.

[CM06] N. Chernov & R. Markarian, *Dispersing billiards with cusps: slow decay of correlations*, submitted.

[CY00] N. I. Chernov & L.-S. Young, *Decay of correlations of Lorentz gases and hard balls*. In *Hard Ball Systems and Lorentz Gas*, D. Szász, editor, Springer, Berlin (2000), 89–120.

[CZ05] N. Chernov & H.-K. Zhang, *Billiards with polynomial mixing rates*, Nonlinearity **18** (2005), 1527–1553.

[Den89] M. Denker, *The central limit theorem for dynamical systems*, Dyn. Syst. Ergod. Th. Banach Center Publ., **23**, PWN–Polish Sci. Publ., Warsaw, 1989.

[Dev89] R. Devaney, *An introduction to chaotic dynamical systems*, 2nd Ed., Addison-Wesley, New York, 1989.

[DM01] G. Del Magno, *Ergodicity of a class of truncated elliptical billiards*, Nonlinearity **14** (2001), 1761–1786.

[DMM03] G. Del Magno & R. Markarian, *Bernoulli elliptical stadia*, Comm. Math. Phys. **233** (2003), 211–230.

[DMM06] G. Del Magno & R. Markarian, *On the Bernoulli property of planar hyperbolic billiards*, preprint (2006).

[Dob68a] R. L. Dobrushin, *The description of a random field by means of conditional probabilities and conditions of its regularity*, Th. Prob. Appl. **13** (1968), 197–224.

[Dob68b] R. L. Dobrushin, *Gibbsian random fields for lattice systems with pairwise interaction*, Funct. Anal. Appl. **2** (1968), 292–301.

[Dob68c] R. L. Dobrushin, *The problem of uniqueness of a Gibbsian random field and the problem of phase transitions*, Funct. Anal. Appl. **2** (1968), 302–312.

[Don91] V. Donnay, *Using integrability to produce chaos: billiards with positive entropy*, Comm. Math. Phys. **141** (1991), 225–257.

[DPS86] *Dependence in probability and statistics. A survey of recent results*, Eds. E. Eberlein and M. Taqqu, Progress in Probability and Statistics, **11**, Birkhäuser, Boston, MA, 1986.

[DS00] *Dynamical Systems, Ergodic Theory and Applications*, Encycl. Math. Sciences, **100**, Ed., Ya. G. Sinai, Springer, Berlin, 2000.

[FM88] B. Friedman & R. F. Martin, *Behavior of the velocity autocorrelation function for the periodic Lorentz gas*, Physica D **30** (1988), 219–227.

[Ga74] G. Gallavotti, *Lectures on the billiard*. In *Dynamical systems, theory and applications (Rencontres, Battelle Res. Inst., Seattle, Wash., 1974)*, Lect. Notes Phys. **38**, Springer, Berlin (1975), 236–295.

[GG94] P. Garrido & G. Gallavotti, *Billiards correlation function*, J. Stat. Phys. **76** (1994), 549–585.

[GKT95] G. Galperin, T. Krüger, & S. Troubetzkoy, *Local instability of orbits in polygonal and polyhedral billiards*, Comm. Math. Phys. **169** (1995), 463–473.

[GO74] G. Gallavotti & D. S. Ornstein, *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83-101.

[Gur65] B. M. Gurevich, *Construction of increasing partitions for special flows*, Th. Prob. Appl. **10** (1965), 627–645.

[Gut86] E. Gutkin, *Billiards in polygons*, Physica D **19** (1986), 311–333.

[Gut96] E. Gutkin, *Billiards in polygons: survey of recent results*, J. Stat. Phys. **83** (1996), 7–26.

[GZ90] G. A. Galperin & A. N. Zemlyakov, *Mathematical billiards*, Kvant **77**, Nauka, Moscow, 1990.

[Had01] J. Hadamard, *Sur l'itération et les solutions asymptotiques des équations différentielles*, Bull. Soc. Math. France **29** (1901), 224–228.

[Hal77] B. Halpern, *Strange billiard tables*, Trans. AMS **232** (1977), 297–305.

[HB00] *Hard Ball Systems and Lorentz Gas*, D. Szász, editor, Springer, Berlin, 2000.

[Ho39] E. Hopf, *Statistik der geodetischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261–304.

[Ho40] E. Hopf, *Statistik der Lösungen geodätischer Probleme vom unstabilen Typus, II*, Math. Annalen **117** (1940), 590–608.

[HP70] M. Hirsch & C. Pugh, *Stable manifolds and hyperbolic sets*. In: Global Analysis, Proc. Symp. in Pure Math., **14**, AMS, Providence, RI, 1970, pp. 133–163.

[Hu01] H. Hu, *Statistical properties of some almost hyperbolic systems* In: Smooth Ergodic Theory and Its Applications (Seattle, WA, 1999), 367–384, Proc. Sympos. Pure Math., **69**, Amer. Math. Soc., Providence, RI, 2001.

[IL71] I. A. Ibragimov & Y. V. Linnik, *Independent and stationary sequences of random variables*, Wolters-Noordhoff, Gröningen, 1971.

[KB94] A. Katok & K. Burns, *Infinitesimal Lyapunov functions, invariant cone families and stochastic properties of smooth dynamical systems*, Ergod. Th. Dynam. Syst. **14** (1994), 757–785.

[KH95] A. Katok & B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge U. Press (1995).

[KM81] I. Kubo & H. Murata, *Perturbed billiard systems II, Bernoulli properties*, Nagoya Math. J. **81** (1981), 1–25.

[KMS86] S. Kerckhoff, H. Masur & J. Smillie, *Ergodicity of billiard flows and quadratic differentials*, Ann. Math. **124** (1986), 293–311.

[KS86] A. Katok & J.-M. Strelcyn, with the collaboration of F. Ledrappier & F. Przytycki, *Invariant manifolds, entropy and billiards; smooth maps with singularities*, Lect. Notes Math., **1222**, Springer, New York (1986).

[KSS90] A. Krámli, N. Simányi & D. Szász, *A "transversal" fundamental theorem for semi-dispersing billiards*, Comm. Math. Phys. **129** (1990), 535–560

[KSS91] A. Krámli, N. Simányi & D. Szász, *The K–property of three billiard balls*, Ann. Math. **133** (1991), 37–72

[KSS92] A. Krámli, N. Simányi & D. Szász, *The K-property of four billiard balls*, Comm. Math. Phys. **144** (1992), 107–142.

[KT91] V. V. Kozlov & D. V. Treshchëv, *Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts* Translations of Mathematical Monographs, **89**, AMS, Providence, RI (1991).

[La73] V. F. Lazutkin, *On the existence of caustics for the billiard ball problem in a convex domain*, Math. USSR Izv. **7** (1973), 185–215.

[Led84] F. Ledrappier, *Propriétés ergodiques des mesures de Sinai*, IHES Public. Math., **59** (1984), 163–188.

[Leo61] V. P. Leonov, *On the dispersion of time-dependent means of a stationary stochastic process*, Th. Probab. Appl., **6** (1961), 87–93.

[Ll92] R. de la Llave, *Smooth conjugacy and S-R-B measures for uniformly and non-uniformly hyperbolic systems*, Comm. Math. Phys. **150** (1992), 289–320.

[LMM86] R. de la Llave, J. Marco & R. Moriyón, *Canonical perturbation theory of Anosov systems and regularity results for the Livsic cohomology equation*, Ann. Math. **123** (1986), 537–611.

[Lo05] H. A. Lorentz, *The motion of electrons in metallic bodies*, Proc. Amst. Acad. **7** (1905), 438–453.

[LR69] O. E. Lanford & D. Ruelle, *Observables at infinity and states with short range correlations in statistical mechanics*, Comm. Math. Phys. **13** (1969), 194–215.

[LS82] F. Ledrappier & J.-M. Strelcyn, *A proof of the estimation from below in Pesin's entropy formula*, Erg. Th. Dynam. Syst. **2** (1982), 203–219.

[LW95] C. Liverani & M. Wojtkowski, *Ergodicity in Hamiltonian systems*, Dynamics reported, Dynam. Report. Expositions Dynam. Systems (N.S.) **4**, Springer, Berlin (1995), 130–202.

[Mac83] J. Machta, *Power law decay of correlations in a billiard problem*, J. Stat. Phys. **32** (1983), 555–564.

[Man83] R. Mañe, *Introducao à Teoria Ergódica*. IMPA, Rio de Janeiro (1983). English Ed., *Ergodic Theory and Differentiable Dynamics*, Springer, Berlin, Heidelberg, New York (1987).

[Mar88] R. Markarian, *Billiards with Pesin region of measure one*, Comm. Math. Phys. **118** (1988), 87–97.

[Mar93a] R. Markarian, *The fundamental theorem of Sinai-Chernov for dynamical systems with singularities*. In *Dynamical Systems. Santiago de Chile 1990*, R. Bamón, R. Labarca, J. Lewowicz, and J. Palis, editors, Longman, Harlow (1993), 131–158.

[Mar93b] R. Markarian, *New ergodic billiards: exact results*, Nonlinearity **6** (1993), 819–841.

[Mar94] R. Markarian, *Non-uniformly hyperbolic billiards*, Ann. Fac. Sci. Toulouse Math. **3** (1994), 223–257.

[Mar04] R. Markarian, *Billiards with polynomial decay of correlations*, Erg. Th. Dynam. Syst. **24** (2004), 177–197.

[Mat75] G. Matheron, *Random sets and integral geometry*, J. Wiley & Sons, New York, 1975.

[ME81] N. Martin & J. England, *Mathematical theory of entropy*, Addison-Wesley, Reading, Mass., 1981.

[MOP96] R. Markarian, S. Oliffson & S. Pinto, *Chaotic properties of the elliptical stadium*, Comm. Math. Phys. **174** (1996), 661–679.

[Os68] V. I. Oseledec, *A multiplicative ergodic theorem*, Trans. Moskow Math. Soc. **19** (1968), 197–231.

[OW98] D. S. Ornstein & B. Weiss, *On the Bernoulli nature of systems with some hyperbolic structure*, Erg. Th. Dynam. Syst. **18** (1998), 441–456.

[Pes76] Ya. B. Pesin, *Families of invariant manifolds corresponding to nonzero characteristic exponents*, Math. USSR Izvest. **10** (1976), 1261–1305.

[Pes77a] Ya. B. Pesin, *Characteristic Lyapunov exponents and smooth ergodic theory*, Russ. Math. Surv. **32:4** (1977), 55–114.

[Pes77b] Ya. B. Pesin, *Description of π-partition of a diffeomorphism with invariant measure*, Math. Notes **22** (1977), 506–515.

[Pes92] Ya. B. Pesin, *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, Ergod. Th. Dynam. Sys. **12** (1992), 123–151.

[PeS82] Ya. B. Pesin & Ya. G. Sinai, *Gibbs measures for partially hyperbolic attractors*, Erg. Th. Dynam. Sys. **2** (1982), 417–438.

[Pet83] K. Petersen, *Ergodic Theory*, Cambridge U. Press, 1983.

[PhS75] W. Philipp & W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Memoir. AMS **161** (1975).

[PP90] W. Parry & M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187–188** (1990), 268 pp.

[Re95] J. Rehácek, *On the ergodicity of dispersing billiards*, Rand. Comput. Dynam. **3** (1995), 35–55.

[Ro49] V. Rohlin, *On the fundamental ideas of measure theory*, AMS Transl. Ser. 1, **10**, 1–52, AMS, Providence, RI, 1952.

[Ro67] V. Rohlin, *Lectures on the entropy theory of transformations with invariant measure*, Uspehi Mat. Nauk **22** (1967), 3–56.

[Ru76] D. Ruelle, *A measure associated with Axiom-A attractors*, Amer. J. Math. **98** (1976), 619–654.

[Ru78] D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, Reading, Mass., 1978.

[Ru79] D. Ruelle, *Ergodic theory of differentiable dynamical systems*, Publ. Math. IHES **50** (1979), 27–58.

[Sa76] L. A. Santaló, *Integral Geometry and Geometric Probability*, Addison Wesley, Reading, Mass., 1976.

[SC87] Ya. G. Sinai & N. I. Chernov, *Ergodic properties of some systems of 2-dimensional discs and 3-dimensional spheres*, Russ. Math. Surv. **42** (1987), 181–207.

[Sim99] N. Simányi, *Ergodicity of hard spheres in a box*, Ergod. Th. Dynam. Syst. **19** (1999), 741–766.

[Sim03] N. Simányi, *Proof of the Boltzmann-Sinai ergodic hypothesis for typical hard disk systems*, Invent. Math. **154** (2003), 123–178.

[Sim04] N. Simányi, *Proof of the ergodic hypothesis for typical hard ball systems*, Ann. H. Poincaré **5** (2004), 203–233.

[Sin66] Ya. G. Sinai, *Classical dynamic systems with countably-multiple Lebesgue spectrum. II*, Izv. Akad. Nauk SSSR Ser. Mat. **30** (1966), 15–68; also AMS Trans. **68** (1968), 34–88.

[Sin68a] Ya. G. Sinai, *Markov partitions and C-diffeomorphisms*, Funct. Anal. Appl. **2** (1968), 61–82.

[Sin68b] Ya. G. Sinai, *Construction of Markov partitions*, Funct. Anal. Appl. **2** (1968), 245–253.

[Sin70] Ya. G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.

[Sin72] Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surv. **27** (1972), 21–69.

[Sin76] Ya. G. Sinai, *Introduction to Ergodic Theory*. Princeton U. Press, Princeton (1976).

[Sin79] Ya. G. Sinai, *Development of Krylov's ideas*, Afterword to N. S. Krylov, *Works on the Foundations of Statistical Physics*. Princeton U. Press, Princeton, NJ, (1979), 239–281.

[St86] J.-M. Strelcyn, *Plane billiards as smooth dynamical systems with singularities*, In: Lect. Notes Math., **1222**, Springer, New York (1986), pp. 199–278.

[SV04] D. Szász & T. Varjú, *Markov towers and stochastic properties of billiards*. In: Modern Dynamical Systems and Applications, 433–445, Cambridge U. Press, Cambridge, 2004.

[Sz92] D. Szász, *On the K-property of some planar hyperbolic billiards*, Comm. Math. Phys. **145** (1992), 595–604.

[Ta95] S. Tabachnikov, *Billiards*, Panor. Synth. No. 1, SMF, Paris (1995).

[Va92] S. Vaienti, *Ergodic properties of the discontinuous sawtooth map*, J. Stat. Phys. **67** (1992), 251–269.

[Wa82] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York (1982).

[Wo85] M. Wojtkowski, *Invariant families of cones and Lyapunov exponents*, Ergod. Th. Dynam. Syst. **5** (1985), 145–161.

[Wo86] M. Wojtkowski, *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Comm. Math. Phys. **105** (1986), 391–414.

[Wo94] M. Wojtkowski, *Two applications of Jacobi fields to the billiard ball problem*, J. Differ. Geom. **40** (1994), 155–164.

[Yo98] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. Math. **147** (1998), 585–650.

[Yo99] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188.

[ZE97] G. M. Zaslavsy & M. Edelman, *Maxwell's demon as a dynamical system*, Phys. Rev. E **56** (1997), 5310–5320.

Index