Mathematical
Surveys
and
Monographs

Volume 127

Chaotic Billiards

Nikolai Chernov
Roberto Markarian

American Mathematical Society

Mathematical
Surveys
and
Monographs

Volume 127

Chaotic Billiards

Nikolai Chernov
Roberto Markarian

American Mathematical Society

2000 *Mathematics Subject Classification.* Primary 37D50; Secondary 37D25, 37A25, 37N05, 82B99.

For additional information and updates on this book, visit
www.ams.org/bookpages/surv-127

Library of Congress Cataloging-in-Publication Data

Chernov, Nikolai, 1956–.
Chaotic billiards / Nikolai Chernov, Roberto Markarian.
p. cm. — (Mathematical surveys and monographs, ISSN 0076-5376 ; v. 127)
Includes bibliographical references and index.
ISBN 0-8218-4096-7 (alk. paper)
1. Dynamics. 2. Chaotic behavior in systems. 3. Ergodic theory. 4. Measure theory. 5. Probability theory. 6. Billiards. I. Markarian, Roberto. II. Title. III. Series: Mathematical surveys and monographs ; no. 127.

QA851 .C44 2006
003′.857—dc22 2006042819

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/

10 9 8 7 6 5 4 3 2 1 11 10 09 08 07 06

To Yakov Sinai on the occasion of his 70th birthday

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

Contents

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.

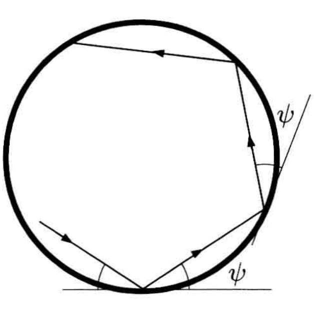

FIGURE 1.1. Billiard motion in a circle.

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.

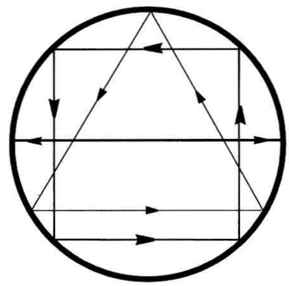

FIGURE 1.2. Periodic motion in a circle.

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

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

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*

$$\theta_n = \theta_0 + 2n\psi_0 \qquad (\text{mod } 2\pi)$$
$$\psi_n = \psi_0.$$

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.

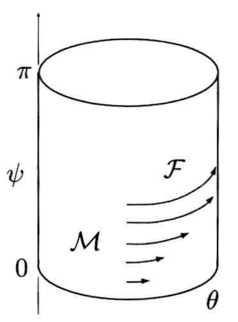

FIGURE 1.3. Action of the collision map $\mathcal{F}$ on $\mathcal{M}$.

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 $\{\theta + 2n\psi,\ \ n \in \mathbb{Z}\}$ 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).

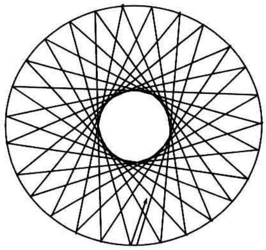

FIGURE 1.4. A nonperiodic trajectory.

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.)

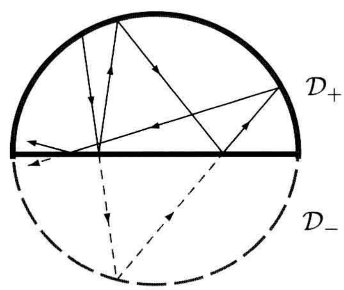

FIGURE 1.5. Billiard in the upper half circle.

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 $\mathcal{D} = \{(x, y) : 0 \le x, y \le 1\}$; 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