Problems on Billiards

Eugene Gutkin *

December 21, 2001

Abstract

This is a discussion of selected open questions in billiard dynamics.

Introduction

Billiard dynamics broadly understood is the geodesic flow on a Riemannian manifold with a boundary. Sometimes even this very general framework is not broad enough. For instance, certain problems in physics lead to the Finsler billiard. This means that the manifold in question is endowed with a Finsler metric. See [33] for an introduction into the subject of the Finsler billiard. Even worse, physical applications lead to the billiard on manifolds with singularities. They are the configuration spaces of the underlying physical systems. In some cases the singular configuration space fits into the framework of *manifolds with corners*. The simplest such manifolds are the planar polygons, and there are physical models that yield triangular billiards. See [16, 34, 22] and the survey [26].

The configuration space of the famous physical system of round elastic balls [67] has a boundary with complicated singularities. It is structured combinatorially like a polyhedron with a large number of faces of all dimensions. But it is much more complicated than a polyhedron, because its faces are not

*Caltech and Max-Planck-Institut for Mathematics in the Sciences (MIS), Inselstrasse 22-26, D 04103 Leipzig, Germany. *E-mail*: gutkin@mis.mpg.de,gutkine@aol.com

flat. The mathematical investigation of this system produced the celebrated Boltzmann Ergodic Hypothesis. After Sinai's seminal paper [62], a modified version of the original conjecture became known as the Boltzmann-Sinai Ergodic Hypothesis. See [62] and the Appendix in [67] for the background, and § 2 for further details.

However, in this exposition we restrict ourselves to the billiard in a bounded planar domain with piecewise smooth boundary. The reason is threefold. First, we concur with the opinion of G. D. Birkhoff [4] that this setting allows to pursue right away interesting qualitative mathematical questions, bypassing lengthy preliminaries and/or cumbersome formalism. Second, there are meaningful physical models that lead to planar billiards. We refer to [68] and to the survey [26] for elaborations. The third, and the most important reason is that the planar billiard offers intriguing open problems. The problems are basic in that they are simple to formulate, and that they concern the basic characteristics of these dynamical systems. Many of these problems deal with the classical subject of periodic billiard orbits.

In the body of the paper we discuss several open problems of billiard dynamics. The choice of the problems is motivated partly by the personal taste, and partly by the simplicity of formulation. To minimize the preliminaries, we refer the reader to [68] for a modern introduction into the billiard ball problem. See also [48], especially for the background on the material of § 1. For obvious reasons, we call the planar domain in question the *billiard table.* The geometric shape of the billiard table determines the qualitative character of the corresponding motion. Historically, three geometric classes of shapes have mostly attracted the mathematicians' attention. First, it is the smooth and strictly convex billard tables. For several reasons, the corresponding billiard dynamics is called elliptic. This is the subject of § 1. Second, we consider the piecewise concave and piecewise smooth billiard tables. The corresponding dynamics is hyperbolic. However, there are also (nonstrictly) convex billiard tables that yield hyperbolic dynamics. See § 2. Finally, billiard tables of the third class are the polygons. The corresponding dynamics is parabolic. See § 3 for a discussion of open problems for polygonal billiards. Some of these problems are completely elementary.

Before turning to the body of the paper, I gratefully acknowledge the remarks of several people on a preliminary draft of this survey. In particular, A. Katok pointed out that he wrote a somewhat similar in structure paper on the "billiard ball problem", based on his lecture at the Independent

University of Moscow [41].

We introduce now the basic notation and the terminology. It will be used throughout the paper. We refer to [68] (pp. 3-4) for the figures illustrating our conventions. Let $Y \subset \mathbf{R^2}$ be a (closed) billiard table. Its boundary ∂Y is a finite union of C^1 curves. We are not assuming that Y is simply connected, thus ∂Y may have several connected components. The *billiard flow* in Y is modeled on the motion of a material point. We will refer to it as the "particle" or the "billiard ball". At each moment of time the state of the system is determined by the position of the ball, $y \in Y$, and its velocity, $v \in \mathbf{R^2}$. It suffices to restrict the attention to the motion with the unit speed. Thus, v is a unit vector. When $y \in \partial Y$, then v is subject to another restriction: It is directed inward. Let Ψ be the space of pairs (y, v), subject to these restrictions. Then Ψ is the *phase space of the billiard flow*. It is three-dimensional. If Y is simply connected, and ∂Y is C^1, then Ψ is homeomorphic to the three-dimensional sphere, but we will not use this observation. If $(y, v) \in \Psi$, we will say that y is the foot-point and v is the direction of the billiard ball. The ball rolls with the unit speed along the straight line through y, in direction v, until it reaches ∂Y. At this instant the direction of the ball changes. Let $y' \in \partial Y$ be the point where the ball touches the boundary of the billiard table. The transformation, $v \mapsto v'$, is the *orthogonal reflection* about the tangent line to ∂Y at y'. Then $(y', v') \in \Psi$, and the ball keeps rolling. These rules define the billiard flow $T^t : \Psi \to \Psi$.

A few remarks are in order. The rules defining the flow T^t stem from the assumptions that the billiard motion is frictionless, and that the boundary of the billiard table is perfectly elastic. The orthogonal reflection rule insures that billiard orbits are the local minimizers of the distance functional. This property was used in [33] to define the reflection rule of the Finsler billiard. If ∂Y does not have a tangent at y', then the transformation $(y, v) \mapsto (y', v')$ is not defined. These are the corner points or the cusp points of the boundary. The standard convention is to "stop the ball" when it reaches a corner point. Thus, if Y is not C^1, then the billiard flow has orbits which are not defined for all times. Our assumptions on Y imply that the union of these orbits has zero volume with respect to the natural invariant measure.

Set $X = \partial Y$, and endow it with the positive orientation. Choosing a reference point on each connected component, and using the arclength as parameter, we identify X with the disjoint union of $k = k(Y)$ circles. In the body of the paper (with a possible exception of § 3) $k = 1$. Denote by $\Phi \subset \Psi$

the subset given by the condition $y \in \partial Y$. Then Φ is a cross-section for the billiard flow. The corresponding Poincaré mapping is the *billiard map*. The terminology is due to G. D. Birkhoff who championed the "billiard ball problem". See [4] and § 1 below. We will use explicit coordinates on Φ. Let x be the arclength parameter on X. For $(x, v) \in \Phi$ let θ be the angle between v and the positive tangent to ∂Y at x. Then $0 \leq \theta \leq \pi$, where 0 and π correspond to the forward and the backward tangential directions respectively. The arclength and the angle are independent coordinates. They yield an isomorphism $\Phi = X \times [0, \pi]$. Since X is a disjoint union of circles, Φ is a disjoint union of k finite cylinders, or annuli. Note that this coordinatisation fails at the corner points of ∂Y. We use the notation $\phi(x, \theta) = (x_1, \theta_1)$ for the billiard map.

The Liouville measure on Ψ is invariant under the billiard flow. Let p, q be the euclidean coordinates in $\mathbf{R^2}$, and let $0 \leq \alpha < 2\pi$ be the angle coordinate on the unit circle. The density of the Liouville measure is $d\nu = dpdqd\alpha$. We denote by μ the *induced measure* on Φ. It is invariant under the billiard map, and we have $d\mu = \sin\theta dx d\theta$. Slightly abusing the language, we will call μ the Liouville measure for the billiard map. Both measures are finite, but not normalized. A straightforward computation yields

$$\nu(\Psi) = 2\pi \text{Area}(Y), \;\; \mu(\Phi) = 2\,\text{Length}(\partial Y). \tag{1}$$

1 Smooth, strictly convex billiards: elliptic dynamics

The first deep investigation of this billiard is due to G. D. Birkhoff. See [4], vol. 2. For this reason, it is often called in the modern literature the *Birkhoff billiard*. The billiard map is an *area preserving twist map*. See, e. g., [43] for more information about these maps, in general, and the Birkhoff billiard, in particular. An *invariant circle* is a ϕ-invariant curve $\Gamma \subset \Phi$ which is a noncontractible topological circle. Recall that the space Φ is a topological annulus. Both components of $\partial\Phi$ are trivial invariant circles. Interpreting Φ as the space of rays intersecting the billiard table, we view Γ as a curve in the space of rays. With any such curve we associate its set of focusing points, $F(\Gamma) \subset \mathbf{R^2}$. Ignoring for the moment possible singularities of $F(\Gamma)$, we point out that for a general curve $\Gamma \subset \Phi$, its set of focusing points is not contained in Y. This is true even for the general invariant curve. For instance, let Y be an ellipse. Let $f, f' \in Y$ be its foci. There are invariant curves Γ such that $F(\Gamma)$ is a confocal hyperbola intersecting the segment $[ff']$.

But if Γ is an invariant circle, then $\gamma = F(\Gamma)$ is contained in Y. Moreover, if Γ is a nontrivial invariant circle, then $\gamma \subset \mathrm{Int}(Y)$. See [31]. This observation extends to the "Birkhoff Minkowski billiard" [33]. The curves γ formed by the focusing points of invariant circles are the *caustics* of the billiard table. In the example above, the caustics are the confocal ellipses contained in Y. Their union is the region $Y \setminus [ff']$. The invariant circles in Φ also fill out a region, $C(\Phi)$, with nonempty interior. Assume that Y is not a disc. Then the complement $\Phi \setminus C(\Phi)$ has a nonempty interior as well. The region $\Phi \setminus C(\Phi)$ looks like a pair of "eyes" in the middle of the cylinder Φ. See [68]. There are many open questions about the caustics. We will formulate only one, the most famous.

Definition 1 *A billiard table Y is* integrable *if the union of invariant circles in the phase space has nonempty interior.*

The following statement is commonly called "the Birkhoff conjecture", although it first appeared in print in a paper by Poritsky [57].

Problem 1. Ellipses are the only integrable billiard tables.

Let Y be a euclidean disc. We may view Y as a degenerate ellipse, with $f = f'$. The preceding analysis applies, but now the invariant circles fill out all of the phase space. M. Bialy proved the converse: If all of Φ is foliated by invariant circles, then Y is a disc [3]. If ∂Y is sufficiently smooth, and its curvature is strictly positive, then the invariant circles in Φ fill out a set of positive measure. This was first proved by V. Lazutkin under the assumption that ∂Y was of class C^{333} [49]. The smoothness requirement was eventially lowered to C^6. See [68] and the references there. By a theorem of J. Mather [54], the positive curvature condition is necessary for the existence of caustics.

An invariant region, $\Omega \subset \Phi$, which is a noncontractible topological annulus whose interior contains no invariant circles, is a *Birkhoff instability region* (or zone). This is an important concept for area preserving twist maps. See [43]. Assume the Birkhoff conjecture, and let Y be a non-elliptical billiard table. Then Φ contains at least one Birkhoff instability zone. An instability region has positive *topological entropy* [1]. Hence, the Birkhoff conjecture implies that any non-elliptical billiard has positive topological entropy. By the *metric entropy* of a billiard we mean its entropy with respect to the Liouville measure. The only examples of convex billiard tables with positive entropy are the Bunimovich stadium and its generalizations. Although these regions are convex, the corresponding billiards are hyperbolic. See § 2. This leads to our next open problem.

Problem 2. Give an example of an elliptic billiard with positive metric entropy.

Since the notion of "elliptic billiard" is somewhat vague, we give a concrete, geometric version of this problem. Note that the Bunimovich stadium and its generalizations are C^1-smooth, but not C^2-smooth. Besides, they are not strictly convex.

Problem 3. a) Construct a strictly convex C^1-smooth billiard table with positive metric entropy. b) Construct a convex C^2-smooth billiard table with positive metric entropy.

Using a simple variational principle, Birkhoff proved the existence of certain periodic billiard orbits [4]. His approach extends to area preserving twist maps, and yields the same existence result [43]. However, the billiard framework makes the considerations especially geometric and elementary. A periodic orbit of period q (under the billiard map) corresponds to a (directed) closed polygon, P, with q sides, inscribed in ∂Y. Vice versa, every closed (directed) inscribed q-gon, satisfying the obvious condition on the angles it makes with ∂Y, determines a periodic orbit of period q. Birkhoff called polygons satisfying the angles condition the *harmonic polygons*. Let $1 \leq p < q$ be the number of times the pencil tracing P in the positive direction goes around ∂Y. The ratio $0 < p/q < 1$ is the *rotation number* of a periodic orbit. Fix a pair $1 \leq p < q$, with p and q relatively prime. Let $X(p, q)$ be the set of all incribed q-gons that go p times around ∂Y. The space $X(p, q)$ is a manifold with corners. For $P \in X(p, q)$ set $f(P)$ be the physical circonference of P. Then the harmonic polygons are the critical points of the function f. Birkhoff proved that f has at least two distinct critical points. One of them delivers the maximum, and the other a minimax to the length function. The corresponding periodic billiard orbits are the *Birkhoff periodic orbits* with rotation number p/q.

As an example, let us consider the rotation number $1/2$. Then the maximal Birkhoff orbit yields the diameter of Y. The minimax orbit corresponds to the width of Y. When the two are equal, ∂Y is a *curve of constant width*, and we have a one-parameter family of periodic orbits with rotation number $1/2$. They fill out the "equator" of Φ. There are other examples of the Birkhoff billiard tables with one-parameter families of periodic orbits having the same length and the same rotation number. See [37] and [25] for different approaches.

One of the basic characteristics of a dynamical system is the *growth rate of periodic points*. In order to talk about a growth rate, we need to introduce a way to account for the number of these points. These accounting devices are usually called the *counting functions*. There are different kinds of counting functions. One of them is the number of periodic points of the period at most n. We mean the period with respect to the billiard map. Let $f_Y(n)$ denote this counting function. (See §§ 2 and 3 for other examples.) Denote by $g_Y(n)$ the number of periodic orbits of period at most n.

Birkhoff's theorem yields a universal quadratic lower bound on $g_Y(n)$. Namely, $g_Y(n)/2$ is bounded below by the number of relatively prime pairs $1 \leq p < q \leq n$. This implies a universal cubic lower bound on the counting function $f_Y(n)$, i. e., $f_Y(n) \geq cn^3$. We leave the computation of the universal constant c in this inequality to the reader. See, e. g., [36].

In view of the examples above, $f_Y(n)$ may be infinite. Hence, there is no universal upper bound on the number of periodic points. Another natural way to estimate the size of a set is to compute its measure. Let Φ be the phase space of the billiard map, and let $\mathcal{P} \subset \Phi$ (resp. $\mathcal{P}_n \subset \Phi$) be the set of periodic points (resp. periodic points of period n). For example, if Y is a table of constant width, then the set $\mathcal{P}_2$ is the equator of Φ. Although it is infinite, $\mu(\mathcal{P}_2) = 0$. We formulate our next open problem as a claim.

Problem 4. Prove that for any Birkhoff billiard table $\mu(\mathcal{P}) = 0$.

Since $\mathcal{P} = \cup_{n=2}^{\infty} \mathcal{P}_n$, a disjoint union, the preceding claim is equivalent to the following.

Problem 5. Prove that for any Birkhoff billiard table $\mu(\mathcal{P}_n) = 0$ for all n.

The natural analog of this claim for the general area preserving twist maps fails! This problem is strictly about the billiard. For $n = 2$ the solution is straightforward. In fact, we have already outlined the argument in the preceding discussion. For $n = 3$ this is a theorem of M. Rychlik [59]. Rychlik's proof depends on a formal identity, which he verified using Maple. L. Stojanov simplified the proof, and eliminated the computer verification [65]. Ya. Vorobets gave an independent proof [72]. His argument applies to higher dimensional billiards as well. Finally, M. Wojtkowski [76] proved the claim, as an application of the *mirror equation* of the geometric optics. See [31] for other applications of the geometric optics in billiard dynamics.

For $n \geq 4$ the question is open. To obtain a counterexample, one would have to produce a billiard table with a two-parameter family of periodic orbits. Recall that there are billiard tables with one-parameter families of such orbits [37, 25]. However, these tables are very special. In fact, the claim of Problem 4 should hold for any billiard table. It has nothing to do with the ellipticity. There is plenty of partial evidence that the claim holds. For instance, this is known for the tables of §§ 2 and 3. It was proved for the billiard tables with piecewise real analytic boundary. See the book [60] and the references there. In addition, a theorem of V. Petkov and L. Stojanov [56] implies that for the generic billiard table (not necessarily Birkhoff!) the sets $\mathcal{P}_n$ are finite for all n.

A solution of Problem 4 would have important implications for analysis. The famous *Weyl conjecture* [74] predicts the second term and the error estimate for the spectral asymptotics of the Laplace operator (with either Dirichlet or Neumann boundary conditions) in a bounded domain. A theorem of V. Ivrii [38] establishes the Weyl conjecture under the assumption that the union of periodic billiard orbits in the domain has measure zero. Thus, a solution of Problem 4 would yield the Weyl conjecture for planar domains.

2 Piecewise concave billiard tables: hyperbolic dynamics

As a geometric motivation, we propose the following construction. Take a convex polygon, P. For instance, P may be a triangle or a quadrilateral.

Then replace each side of the polygon by a circular arc, centered at a distant point. If the center-points are sufficiently far from P, then these circular arcs form a "curvilinear polygon" whose vertices coincide with the vertices of P. Choosing the center-points appropriately, we insure that the "sides" of this curvilinear polygon, Y, are convex inward. This is a particular example of a piecewise concave billiard table. It is not important that the sides of Y are circular. The crucial condition is that they are smooth and convex inward. See the figures in [68], p. 115.

This class of billiard tables arose in the work of Ya. Sinai on the *Boltzmann-Sinai gas* [62]. See the Appendix by D. Szasz in [67] for the background material. In the *Boltzmann gas* the round molecules are confined by the walls of a container. Sinai has replaced the walls by periodic boundary conditions. Thus, the round molecules of the Boltzmann-Sinai gas move on a torus. In the "real world situation", the number of the moving molecules is very large, and the space is three dimensional. Sinai began by considering a "mathematical caricature" of the physical system. In the caricature gas there are only two round molecules, and the space is two dimensional. Eliminating the center of mass, we reduce this system to the geodesic flow on a flat torus with a round hole. Represent the torus by the 2×2 square, so that the hole is the central disc of radius $1/2$. Using the four-fold symmetry of the problem, we reduce it to the billiard in the unit square with the deleted quarter-disc of radius $1/2$, centered at a vertex.

The domain we have constructed is known as *the Sinai billiard* [1]. We will use this example to establish the terminology. Let Y be the domain above. The boundary of Y is a union of four straight segments and a circular arc. The former are the *neutral components* and the latter is a *dispersive component.* Billiard tables like this are called *semi-dispersive.* In the absence of neutral components, we speak of *dispersive* billiard tables. It is the dispersive boundary components that cause the hyperbolicity of the billiard dynamics. We will not give a formal definition of the *hyperbolic billiard dynamics.* We refer the reader to [43] for the background on hyperbolic dynamics, in general, and to § 5 of [68] and to the survey [8] for an introduction into the hyperbolic billiard dynamics, in particular.

[1] Unfortunately, there is a fair amount of confusing terminology in the literature. Mathematicians often use the expressions like "Sinai's billiard" or "the Sinai table" or "a dispersive billiard" interchangeably. Physicists tend to mean by "the Sinai billiard" a special billiard table, though not necessarily the one we just defined.

Slightly abusing the language, we will say that a billiard table, Y, is hyperbolic if the corresponding billiard map, $\phi : \Phi \to \Phi$, is hyperbolic. Already in [62] Sinai proved the hyperbolicity of dispersive tables. After the discovery by L. Bunimovich that the *stadium* and similar billiard tables are hyperbolic [7], the workers in the field started searching for geometric criteria of hyperbolicity. The notion of an *invariant cone field* [75] provided a convenient approach to the subject. See also [42].

We restrict the discussion of this material to billiards. Denote by V_z the tangent plane to the phase space at $z \in \Phi$. The differential ϕ_* is a linear map from V_z to $V_{\phi(z)}$.

Definition 2 *A family $\mathcal{C} = \{C_z \subset V_z : z \in \Phi\}$ is an invariant cone field if the following conditions are satisfied.*

- *1. The closed cone C_z is defined for almost all $z \in \Phi$, and the "function" $z \mapsto C_z$ is measurable.*
- *2. The cone C_z is nontrivial and has nonempty interior.*
- *3. We have $\phi_*(C_z) \subset C_{\phi(z)}$.*
- *4. For almost all $z \in \Phi$ there exists $n = n(z)$ such that $\phi_*^n(C_z) \subset Int(C_{\phi^n(z)})$.*

The existence of an invariant cone field is equivalent to the (nonuniform) hyperbolicity of the billiard [75]. Using a geometric approach, M. Wojtkowski constructed invariant cone fields for several classes of billiard tables [76]. In addition to the "old" classes of hyperbolic tables (i. e., the dispersive tables and the generalized Bunimovich' stadia), Wojtkowski found invariant cone fields for a wide class of locally strictly convex tables. Wojtkowski's principles of the design of hyperbolic billiard tables were further extended by V. Donnay and R. Markarian. See § 5 of [68] and the references there. Recently, the author and his collaborators used Wojtkowski's ideas to construct hyperbolic billiards on surfaces of arbitrary constant curvature [32].

Despite the recent advances in extending the class of hyperbolic billiard tables, the following "old" question remains open.

Problem 6. Is every semi-dispersive billiard table hyperbolic?

We will formulate a more concrete version of this question. Like in the beginning of this section, we start with a (not necessarily convex) n-gon P. We choose $1 \leq m < n$ of the sides, and replace them by convex inward circular arcs of sufficiently large radii. This defines a special class of semi-dispersive billiard tables. We will call them (for want of a better name) the *semi-dispersive polygons*.

Problem 7. Is every semi-dispersive polygon hyperbolic?

When $m = n - 1$, a semi-dispersive polygon, Y, has only one neutral component. Let Y' be the reflection of Y about this side, and set $Z = Y \cup Y'$. Since Z is a dispersive billiard table, it is hyperbolic. Because of the axial symmetry of Z, the billiard maps in Y and Z are, essentially, equivalent. Hence, the table Y is hyperbolic. In special cases, we can extend the trick of reflection, to prove the hyperbolicity of semi-dispersive polygons with $m < n-1$. For instance, let P be a triangle with an angle π/n. Let Y be the semi-dispersive triangle, whose only dispersive component replaces the side of P, opposite the π/n angle. Reflecting Y successively $2n$ times, in an obvious way, we obtain a dispersive billiard table, Z. Thus, Z is hyperbolic. Since Y and Z are related by a $2n$-fold symmetry, the table Y is hyperbolic as well.

A suitable generalization of the reflection trick will work if P is a *rational polygon*. See § 3 below for the definition. The crucial special case of Problem 7, when $m = 1$ and n is arbitrary, is equivalent to Problem 12 of § 3.

Dispersive billiard tables are ergodic [10], while for the hyperbolic tables, in general, this is not the case [76]. However, the stadium and its relatives are ergodic [66]. So far there are no examples of strictly convex hyperbolic billiard tables. See Problems 2 and 3. The billiard dynamics in hyperbolic tables has strong chaotic properties [9, 10, 13]. See also the articles in [67]. Many open questions for hyperbolic billiards have to do with the *decay of correlations*. See, e. g., [14, 79]. Since this subject is rather technical, we do not formulate any of them here. Instead, we will discuss the statistics of periodic orbits in dispersive billiards. In this case, the set of periodic points of a particular period is finite, and the natural counting function is well defined. Let $f_Y(n)$ be the number of k-periodic orbits, with $k \leq n$. The asymptotic behavior of $f_Y(n)$, as $n \to \infty$, is an important dynamical characteristic. From now and until the end of this section, we restrict our discussion to dispersive billiard tables.

By a theorem of Stojanov [64], $f_Y(n) \leq e^{h_+ n}$, for some positive number h_+, as $n \to \infty$. By a theorem of Chernov [9], there is also an exponential lower bound: $e^{h_- n} \leq f_Y(n)$. Combining the two bounds, we obtain the inequalities

$$0 < h_- \leq \liminf_{n\to\infty} \frac{\log f_Y(n)}{n} \leq \limsup_{n\to\infty} \frac{\log f_Y(n)}{n} \leq h_+ < \infty. \tag{2}$$

The following two problems were contributed by N. Chernov [2].

Problem 8. Does the limit below exist?

$$h = \lim_{n\to\infty} \frac{\log f_Y(n)}{n} \tag{3}$$

By equation (2), if this limit exists, then $0 < h_- \leq h \leq h_+ < \infty$.

Problem 9. If the limit in equation (3) does exist, is h the *topological entropy* of the billiard map?

To simplify the exposition, we formulated both questions for dispersive billiards only. Their natural analogs make sense (and are open) for all hyperbolic billiard tables. Moreover, Problems 8 and 9 fit into the general relationship between the distribution of periodic points and the topological entropy [39]. However, the singularities, which is the paramount feature of the billiard dynamics, preclude the applicability of the smooth ergodic theory. Other techniques have to be used. See, for instance, [12] and [28].

3 Polygonal billiards: parabolic dynamics

We refer to [24, 26] and to [68], Chapter 3 for the background material. The surveys [63] and [53] discuss more recent developments. The polygon is not

[2]Personal communication.

required to be convex or even simply connected. We also allow *slits*, i. e., obstacles without interior. A polygon P is *rational* if the angles between the sides of Y are of the form $\pi m/n$. Let $N = N(P)$ be the least common denominator of these rational numbers. A classical construction associates with Y a closed surface $S = S(P)$ tiled by $2N$ copies of P. The surface S has a finite number of cone points, where the total angle is a multiple of 2π. If P is simply connected, then there is a simple formula for the genus of $S(P)$ in terms of the angles of P [24]. This formula implies that $S(P)$ is a torus if and only if P tiles the plane under reflections. The billiard in P is essentially equivalent to the geodesic flow on $S(P)$. This observation was first exploited by A. Katok and A. Zemlyakov [45]. In the papers on billiards $S(P)$ often goes by the name of the "Katok-Zemlyakov surface". However, the construction has been in the literature (at least) since the early 20-th century. See [23] and the references there.

Surfaces $S(P)$ are examples of *translation surfaces*, which are of independent interest [30]. From the viewpoint of the classical analysis, a translation surface is a closed Riemann surface with a holomorphic linear differential. Considering the quadratic differentials instead, we come to the class of *half-translation surfaces*. The connection between the polygonal billiard and the classical complex analysis proved to be very useful for the former [47, 50, 51, 70]. We will give a few examples that show the usefulness, but also the limitations of this relationship.

The geodesic flow of any translation surface, S, decomposes into the one-parameter family of *directional flows* b_θ^t, $0 \leq \theta < 2\pi$. The flow b_θ^t is identified with the linear flow on S in direction θ. The Lebesgue measure on S is preserved by b_θ^t. Hence, the billiard flow in a rational polygon is not ergodic. It decomposes into the one-parameter family of *directional billiard flows*. Let S be an arbitrary translation surface. A theorem of Kerckhoff, Masur, and Smillie [47] says that the flows b_θ^t are uniquely ergodic for Lebesgue almost all θ. As a consequence, the directional billiard flow of a rational polygon is ergodic for almost every direction. For the typical translation surface, the set $\mathcal{N}(S) \subset [0, 2\pi)$ of non-uniquely ergodic directions has positive Hausdorff dimension [52]. Since the typical translation surface does not correspond to a polygon, the theorem does not produce applications to the billiard. However, the corresponding set $\mathcal{N}(P)$ is well understood for special rational polygons [23, 70, 15].

Now we turn to the billiard in irrational (i. e., arbitrary) polygons. Denote by $\mathcal{T}(n)$ the parameter space of euclidean n-gons, up to the scaling. Note that the billiard is preserved by scaling. The space $\mathcal{T}(n)$ is a finite union of natural components corresponding to a particular combinatorial data. Each component is homeomorphic to a relatively compact set of the maximal dimension in a Euclidean space. Thus $\mathcal{T}(n)$ has a natural probability measure, μ, whose restriction to each component is proportional to the corresponding Lebesgue measure. For instance, the space $\mathcal{T}(3)$ of triangles is the subset of $\mathbf{R}^2$, given by $\mathcal{T}(3) = \{(\alpha, \beta) : 0 < \alpha \leq \beta, \alpha, \beta < \pi/2\}$. Thus, $\mathcal{T}(3)$ is, essentially, a square. Since [47], it is known that the set $\mathcal{E}(n) \subset \mathcal{T}(n)$ of ergodic n-gons is residual in the sense of Baire category [55]. A theorem of Ya. Vorobets [73] produced actual examples of ergodic polygons. See more on this below. The following question arises.

Problem 10. Is $\mu(\mathcal{E}(n)) > 0$?

This question is open for all $n \geq 3$. The case of $n = 3$ is especially interesting, since the mechanical system of three elastic particles confined to move on a circle reduces to the billiard in an acute triangle. This was independently noted in [22] and [11]. See [53] for a more detailed account. Let m_1, m_2, m_3 be the masses. Then the angles of the triangle $\Delta(m_1, m_2, m_3)$ satisfy

$$\tan \alpha_i = m_i \sqrt{\frac{m_1 + m_2 + m_3}{m_1 m_2 m_3}}, \ i = 1, 2, 3. \tag{4}$$

Note that in this correspondence between the mechanics and the billiard, the condition of π-rationality of angles does not have an obvious physical meaning. In the limit, when $m_3 \to \infty$, we obtain the physical system of two elastic particles on an interval. The limit of $\Delta(m_1, m_2, m_3)$ is the right triangle whose angles satisfy $\tan \alpha_1 = \sqrt{m_1/m_2}, \tan \alpha_2 = \sqrt{m_2/m_1}$.

Vorobets [73] proved that if the angles of an irrational polygon P admit a particular superexponentially fast rational approximation, then P is ergodic. This result produced examples of ergodic polygons, but it did not solve the problem above. Numerical experiments indicate that irrational polygons are ergodic, and have other stochastic properties [2, 11]. There is no theorem confirming this, but, on the other hand, there is no theorem precluding this possibility. [3]

[3]The following problem grew out of discussions with D. Kleinbock in the Summer of

Problem 11. Give an example of an irrational but nonergodic polygon (triangle).

Another open question has to do with the structure of possible invariant sets in the phase space of a nonergodic polygon. Let P be an arbitrary n-gon. Denote its sides by $a_1, \dots, a_n$. For $1 \le i \le n$ let $\Phi_i \subset \Phi$ be the set of elements in the phase space of the billiard map, whose footpoints belong to the side a_i. Then $\Phi = \cup_{i=1}^n \Phi_i$, a disjoint decomposition. Let μ be the Liouville measure in Φ. Then, by equation (1), $\mu(\Phi_i) = 2\,\mathrm{Length}(a_i)$. Our next open question is motivated by semi-dispersive billiard tables. See Problem 7 in § 2. We formulate this question as a statement.

Problem 12. Let P be an arbitrary n-gon, and let $X \subset \Phi$ be a measurable invariant subset. If X contains one of the sets Φ_i, then $X = \Phi$.

Note that for rational polygons, which are very non-ergodic, the claim holds. This follows from the main result of [47].

Periodic billiard orbits in polygons is a subject that does not require any mathematical background beyond the school geometry, and it has immediate applications to physics. For instance, let $\Delta(m_1, m_2, m_3)$ be the acute triangle corresponding to the system of three elastic masses. See equation (4). Then the periodic billiard orbits in the triangle $\Delta(m_1, m_2, m_3)$ correspond to the periodic trajectories of the physical system of three masses. [4] The reader will appreciate that the problems on periodic billiard orbits in polygons proved to be especially elusive.

Problem 13. Does every polygon have a periodic orbit ?

Every rational polygon has periodic orbits, and much more is known (see below). Certain special classes of irrational polygons have periodic orbits [16, 34]. Every acute triangle has a classical periodic orbit - the Fagnano

1999. However, S. Troubetzkoy informed me that he had posed it several years ago in Oberwolfach.

[4]Another connection between billiards and physics arises in the study of mechanical linkages. See www.ma.huji.ac.il/~drorbn/People/Eldar/thesis for more information on these “Basic Machines”. I am indebted to M. Steiner for bringing this to my attention.

orbit [27]. But it is not known if every acute triangle has other periodic orbits. Neither is it known if every obtuse triangle has a periodic orbit. See [21] and [35] for periodic orbits in obtuse triangles.

Every periodic orbit in a polygon, with an even number of segments, is contained in a *parallel band* of periodic orbits of the same length. The boundary of a band is a union of singular orbits, or the *generalised diagonals* [40]. These are the billiard orbits with endpoints at the vertices. Periodic orbits with an odd number of segments are necessarily isolated, but they are extremely rare. A rational polygon has a finite number of them. Let $f_P(\ell)$ be the number of periodic bands of length at most ℓ. This is the right *counting function* of periodic orbits in polygons.

Problem 14. Find efficient upper and lower bounds on $f_P(\ell)$ for irrational polygons.

At present it is only known that $f_P(\ell)$ grows subexponentially [40, 28]. The concensus is that there should be a universal polynomial upper bound on this counting function. As for a lower bound, it is anybody's guess.

For rational polygons the situation is different. By results of H. Masur [50, 51] and M. Boshernitzan [5, 6], the counting function has quadratic bounds. More precisely, there exist $0 < c_*(P) \leq c^*(P) < \infty$ such that $c_*(P)\ell^2 \leq f_P(\ell) \leq c^*(P)\ell^2$ for sufficiently large ℓ. We can take for $c_*(P)$ and $c^*(P)$ the lower and the upper limits, respectively, of $f_P(\ell)/\ell^2$.

Problem 15. Find efficient estimates on $c_*(P)$ and $c^*(P)$ for rational polygons.

In all examples computed so far $f_P(\ell)/\ell^2$ has a limit, i. e., $c_*(P) = c^*(P) = c(P)$. If this is the case, we say that the *polygon has quadratic asymptotics.* For certain classes of rational polygons this has been proven, and various expressions for the *quadratic constant* $c(P)$ were found. See [23, 69, 71], and [30]. The preceding definitions and questions have their counterparts for translation surfaces, where periodic billiard orbits are replaced by the closed geodesics. On the one hand, the quadratic asymptotics have been established for a special class of polygons and surfaces: Those satisfying the *lattice property*, or, simply, for *lattice polygons and translation surfaces* [69, 29, 30]. The quadratic constants for lattice polygons are number theoretic. See [69, 71] and [30]. They are highly unstable under perturbations. Besides, lattice polygons are rare. All acute lattice triangles have been determined [46, 58]. For obtuse triangles the question is open. On the other hand, it is known that the generic (with respect to the Lebesgue measure) translation surface has quadratic asymptotics. Let S be the generic translation surface. Then the value of its quadratic constant depends only on the connected component of the stratum of the Teichmüller space, to which S belongs [17]. The "generic" quadratic constants have been computed [20, 19]. The result of [17] has no consequences for polygons, since the corresponding subset of the Teichmüller space has measure zero. However, an extension of the preceding technique establishes the quadratic asymptotics and evaluates the generic quadratic constants for a very special but interesting class of rational polygons [18].

Problem 16. Does every rational polygon have quadratic asymptotics?

References

[1] S. Angenent *A remark on the topological entropy and invariant circles of an area preserving twist map*, Proceedings IMA, Vol. 44 (1992), 1 - 6.

[2] R. Artuso, G. Casati, and I. Guarneri, *Numerical study on ergodic properties of triangular billiards*, Phys. Rev. E 55 (1997), 6384 - 6390.

[3] M. Bialy, *Convex billiards and a theorem by E. Hopf*, Math. Zeitschr. 214 (1993), 147 - 154.

[4] G. D. Birkhoff, *Collected Mathematical Papers*, A.M.S., New York, 1950.

[5] M. Boshernitzan, *A condition for minimal interval exchange maps to be uniquely ergodic*, Duke Math. J. 52 (1985), 733 - 752.

[6] M. Boshernitzan, Unpublished Manuscript.

[7] L. Bunimovich, *On the ergodic properties of certain billiards*, Func. Analys. & Appl. 8 (1974), 254 - 255.

[8] L. Bunimovich, *Billiards and other hyperbolic systems*, pp. 192 - 233 in Encyclopedia of Mathematical Sciences vol. 100 (edited by Ya. Sinai), Springer-Verlag, New York, 2000.

[9] L. Bunimovich, Ya. Sinai, and N. Chernov, *Markov partitions for two-dimensional hyperbolic billiards*, Russ. Math. Surv. 45 (1990), 105 - 152.

[10] L. Bunimovich, Ya. Sinai, and N. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Russ. Math. Surv. 46 (1991), 47 - 106.

[11] G. Casati and Prosen, *Mixing property of triangular billiards*, Phys. Rev. Lett. 83 (1999), 4729 - 4732.

[12] N. Chernov, *Topological entropy and periodic points of two-dimensional hyperbolic billiards*, Funct. Anal. Appl. 25 (1991), 39 - 45.

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

[14] N. Chernov, *Decay of correlations in dispersing billiards*, J. Stat. Phys. 94 (1999), 513 - 556.

[15] Y. Cheung, *Hausdorff dimension of the set of nonergodic directions*, preprint, 2001.

[16] B. Cipra, R. M. Hanson, E. Kolan, *Periodic trajectories in right-triangle billiards*, Phys. Rev. **E** 52 (1995), 2066 - 2071.

[17] A. Eskin and H. Masur, *Asymptotic formulas on flat surfaces*, Erg. Theor. & Dyn. Sys. 21 (2001), 443 - 478.

[18] A. Eskin, H. Masur and M. Schmoll, *Billiards in rectangles with barriers*, preprint, 2001.

[19] A. Eskin, H. Masur and A. Zorich, *The Siegel-Veech constants*, in preparation.

[20] A. Eskin and A. Okunkov, *Asymptotics of numbers of branched coverings of a torus and volumes of moduli spaces of holomorphic differentials*, Invent. Math. 145 (2001), 59 - 103.

[21] G. Galperin, A. Stepin and Ya. Vorobets, *Periodic orbits in polygons: generating mechanisms*, Russ. Math. Surv. 47 (1992), 5 - 80.

[22] S. L. Glashow and L. Mittag, *Three rods on a ring and the triangular billiard*, J. Stat. Phys. 87 (1997), 937 - 941.

[23] E. Gutkin, *Billiard flows on almost integrable polyhedral surfaces*, Erg. Theor. & Dyn. Sys. 4 (1984), 560 - 584.

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

[25] E. Gutkin, *Billiard tables of constant width and the dynamical characterizations of the circle*, Workshop on Dynamics and Related Topics, PennState U. (1993), 21 - 24.

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

[27] E. Gutkin, *Two applications of calculus to triangular billiards*, Amer. Math. Monthly 104 (1997), 618 - 623.

[28] E. Gutkin and N. Haydn, *Topological entropy of polygon exchange transformations and polygonal billiards*, Erg. Theor. & Dyn. Sys. 17 (1997), 849 - 867.

[29] E. Gutkin and C. Judge, *The geometry and arithmetic of translation surfaces with applications to polygonal billiards*, Math. Res. Lett. 3 (1996), 391 - 403.

[30] E. Gutkin and C. Judge, *Affine mappings of translation surfaces: geometry and arithmetic*, Duke Math. J. 103 (2000), 191 – 213.

[31] E. Gutkin and A. Katok, *Caustics for inner and outer billiards*, Comm. Math. Phys. 173 (1995), 101 - 133.

[32] B. Gutkin, E. Gutkin and U. Smilansky, *Hyperbolic billiards on surfaces of constant curvature*, Comm. Math. Phys. 208 (1999), 65 - 90.

[33] E. Gutkin and S. Tabachnikov, *Billiards in Finsler and Minkowski geometries*, J. Geom. & Phys. 40 (2001), 277 - 301.

[34] E. Gutkin and S. Troubetzkoy, *Directional flows and strong recurrence for polygonal billiards*, Pittman Res. Notes in Math. N^o 362 (1997), 21 - 45.

[35] L. Halbeisen and N. Hungerbühler, *On periodic billiard trajectories in obtuse triangles*, SIAM Review 42 (2000), 657 - 670.

[36] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford U. Press, London, 1938.

[37] N. Innami, *Convex curves whose points are vertices of billiard triangles*, Kodai Math. J. 11 (1988), 17 - 24.

[38] V. Ivrii, *The second term of the spectral asymptotic expansion for the Laplace-Beltrami operator on manifolds with a boundary*, Func. Anal. & Appl. 14 (1980), 98 - 106.

[39] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Inst. Hautes Etudes Sci. Publ. Math. 51 (1980), 137 - 173.

[40] A. Katok, *The growth rate for the number of singular and periodic orbits for a polygonal billiard*, Comm. Math. Phys. 111 (1987), 151 - 160.

[41] A. Katok, *The billiard table as a mathematician's playground*, (in Russian), Publications of the Moscow Center for Continuing Mathematical Education, 2002, in press.

[42] A. Katok (With the collaboration of Keith Burns), *Infinitesimal Lyapunov functions, invariant cone families and stochastic properties of smooth dynamical systems*, Erg. Theor. & Dyn. Sys. 14 (1994), 757 - 785.

[43] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, CUP, Cambridge, 1995.

[44] A. Katok and J.-M. Strelcyn, *Invariant manifolds, entropy and billiards; smooth maps with singularities*, LNM N^o 1222, Springer-Verlag, Berlin, 1986.

[45] A. Katok and A. Zemlyakov, *Topological transitivity of billiards in polygons*, Math. Notes 18 (1975), 760 – 764.

[46] R. Kenyon and J. Smillie, *Billiards on rational-angled triangles*, Comment. Math. Helv. 75 (2000), 65-108.

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

[48] V. Kozlov and D. Treshchev, *Billiards*, A. M. S., Providence, 1991.

[49] V. Lazutkin, *The existemce of caustics for the billiard problem in a convex domain*, Math. USSR Izv. 7 (1973), 185 - 214.

[50] H. Masur, *Lower bounds for the number of saddle connection and closed trajectories of a quadratic differential*, in "Holomorphic Functions and Moduli", vol. 2, Springer-Verlag, New York, 1988.

[51] H. Masur, *The growth rate of trajectories of a quadratic differential*, Erg. Theor. & Dyn. Sys. 10 (1990), 151-176.

[52] H. Masur and J. Smillie, *Hausdorff dimension of sets of foliations*, Ann. Math. 134 (1991), 455 - 543.

[53] H. Masur and S. Tabachnikov, *Rational billiards and flat structures*, in Handbook of Dynamical Systems, Elsevier, in press.

[54] J. Mather, *Glancing billiards*, Erg. Theor. & Dyn. Sys. 2 (1982), 397 - 403.

[55] J. C. Oxtoby, *Measure and category*, Springer-Verlag, New York, 1996.

[56] V. Petkov and L. Stojanov, *On the number of periodic reflecting rays in generic domains*, Erg. Theor. & Dyn. Sys. 8 (1988), 81 - 91.

[57] H. Poritsky, *The billiard ball problem on a table with a convex boundary - An illustrative dynamical problem*, Ann. Math. 51 (1950), 446 - 470.

[58] J.-Ch. Puchta, *On triangular billiards*, Comment. Mathem. Helv. 76 (2001), 501 – 505.

[59] M. Rychlik, *Periodic points of the billiard ball map in a convex domain*, J. Diff. Geom. 30 (1989), 191 - 205.

[60] Yu. Safarov and D. Vassiliev, *The asymptotic distribution of eigenvalues of partial differential operators*, A.M.S., Providence, 1997.

[61] N. Simanyi, *The K-property of N billiard balls. I*, Inv. Math. 108 (1992), 521 - 548.

[62] Ya. Sinai, *Dynamical systems with elastic reflections: Ergodic properties of dispersing billiards*, Uspekhi Mat. Nauk 25 (1970), 141 - 192.

[63] J. Smillie, *The dynamics of billiard flows in rational polygons*, pp. 360 - 382 in Encyclopedia of Mathematical Sciences vol. 100 (edited by Ya. Sinai), Springer-Verlag, New York, 2000.

[64] L. Stojanov, *An estimate from above on the number of periodic orbits for semi-dispersed billiards*, Comm. Math. Phys. 124 (1989), 217 - 227.

[65] L. Stojanov, *Note on the periodic points of the billiard*, J. Diff Geom. 34 (1991), 835 - 837.

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

[67] D. Szasz (editor), *Hard Ball Systems and the Lorentz Gas*, Encyclopedia of Mathematical Sciences vol. 101, Springer-Verlag, Berlin, 2000.

[68] S. Tabachnikov, *Billiards*, Soc. Math. de France, Paris, 1995.

[69] W. Veech, *Teichmuller curves in modular space, Eisenstein series, and an application to triangulat billiards*, Inv. Math. 97 (1989), 553 - 583.

[70] W. Veech, *Modular spaces of quadratic differentials*, J. Analyse Math. 55 (1990), 117 - 171.

[71] W. Veech, *The billiard in a regular polygon*, GAFA 2 (1992), 341 - 379.

[72] Ya. Vorobets, *On the measure of the set of periodic points of a billiard*, Mat. Notes 55 (1994), 455 - 460.

[73] Ya. Vorobets, *Ergodicity of billiards in polygons*, Mat. Sb. 188 (1997), 65 - 112.

[74] H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, Math. Ann. 71 (1912), 441 - 479.

[75] M. Wojtkowski, *Invariant families of cones and Lyapunov exponents*, Erg. Theor. & Dyn. Sys. 5 (1985), 145 - 161.

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

[77] M. Wojtkowski, *Measure theoretic entropy of the system of hard spheres*, Erg. Theor. & Dyn. Sys. 8 (1988), 133 - 153.

[78] M. Wojtkowski, *Two applications of Jacobi felds to the billiard ball problem*, J. Diff. Geom. 40 (1994), 155 - 164.

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