Journal of Statistical Physics, Vol. 83, Nos. 1/2, 1996

Billiards in Polygons: Survey of Recent Results

Eugene Gutkin[1]

Received January 23, 1995; final May 30, 1995

We review the dynamics of polygonal billiards.

KEY WORDS: Orbit ergodicity; coding; complexity; entropy; periodicity.

1. MOTIVATION AND PRELIMINARIES

In the ten or so years since the publication of ref. 19 polygonal billiards have remained an active subject of research in the mathematics and physics literature. As a result, our understanding of the subject, although still far from being complete, is much better than it was ten years ago. This survey attempts to give a broad overview of the dynamics of billiards in polygons, with an emphasis on the material that was not in the literature in 1985. This is not a survey of the publications on polygonal billiards, and we apologize to authors whose work has not been included. The selection of topics has been strongly influenced by the personal taste of the author, and by space limitations. Thus, we do not discuss quantum polygonal billiards (in particular, quantum chaos; see, e.g., refs. 41, 44, and 45). We hope that somebody will write a survey on this important subject in the near future.

A fascinating aspect of the subject is the interplay between the geometric shape of the billiard table (i.e., a planar curve) and the qualitative features of the billiard dynamics. Hence it is instructive to compare polygonal billiard tables with other classes of billiard tables. In particular, the smooth, strictly convex tables and the dispersing (or Sinai) tables produce strikingly different types of billiard dynamics. We recommend refs. 46 and 15 for a general overview of the subject and a comparative study of billiard dynamics for various types of billiard tables.

---

[1] Mathematics Department, University of Southern California, Los Angeles, California 90089-1113; E-mail: egutkin@math.usc.edu.

0022-4715/96/0400-0007$09.50/0

We assume that the reader is familiar with the basic concepts of planar billiard dynamics in general, and only recall the notions that will be used in what follows. Let P be a closed, connected polygon in the Euclidean plane. We think of P as a bounded region in $\mathbf{R}^2$ whose boundary ∂P consists of a finite number of line segments. It should be clear from the context whether we mean P or ∂P when we talk about a billiard table. We associate with P the billiard flow T^t, $-\infty < t < \infty$, and the billiard map ϕ. The phase space $Z = P \times S^1$ of the flow T^t consists of unit vectors with footpoints in P, and T^t preserves the standard measure, $dx\,dy\,d\theta$. The phase space $\Phi \subset Z$ of ϕ consists of inward-directed vectors with footpoints in ∂P. By construction, Φ is a cross section for T^t, and ϕ: $\Phi \to \Phi$ is the first return map (Fig. 1). The vectors v of Φ are parametrized by (s, θ), where s is the arclength on ∂P and $0 \leqslant \theta \leqslant \pi$ is the angle between v and ∂P. The ϕ-invariant measure on Φ, induced by $dx\,dy\,d\theta$, is $d\mu = \sin\theta\,d\theta\,ds$, and we refer to it as the standard measure on Φ.

The problems in billiard dynamics usually have to do with the behavior of billiard trajectories. The questions mainly fall into two categories: those concerning statistics and those concerning the topology of trajectories. The former belong to ergodic theory, and the latter to topological dynamics. The presentation is organized accordingly. Sections 2 and 3 establish the setting. Section 4 is on ergodic theory, and Section 6 is on the topological dynamics of polygonal billiards. Section 5 is a mixture of results on the borderline between the two kinds of questions.

In the analysis of billiards dynamics, the singularities of orbits produced by the vertices of P play a major role. This is not surprising, since a polygon is essentially determined by its vertices. Surprisingly, the

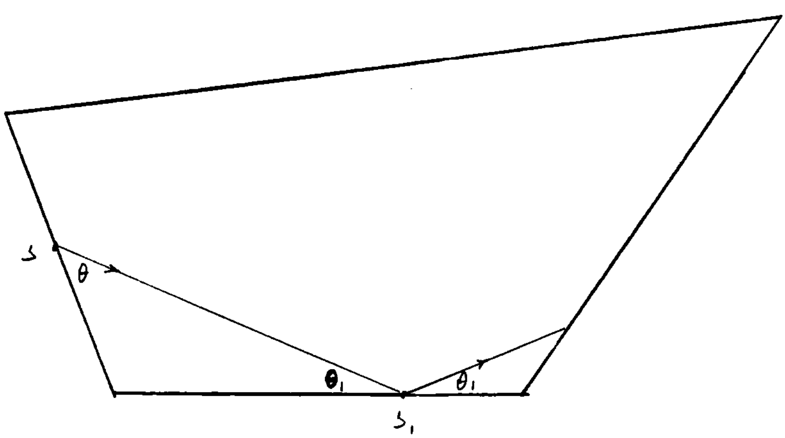

Fig. 1. The billiard map for a polygonal table: $\phi(s, \theta) = (s_1, \theta_1)$.

convexity of P (or the lack of it) does not seem to be of importance. We do not know of any significant property in polygonal billiard dynamics that holds for convex polygons only. The same observation applies to the property of P to be simply connected (i.e., to have no holes inside).

The characteristics of P that are important in the study of billiard dynamics have to do with the vertex angles and the relative lengths of the edges of P. A polygon P is called *rational* if the angles between the lines containing the edges of P are rational multiples of π.[54] Rational polygons play an important role in the research on polygonal billiards.

The methods and tools used in research on polygonal billiards are extremely diverse. They range from Euclidean geometry,[16] to ergodic theory and functional analysis,[29, 33, 5] complex analysis,[31, 34] Lie groups and hyperbolics geometry,[51, 52] general topology[54] and classical analysis.[51]

Consider the mechanical system of two point masses m_1 and m_2 moving freely in the interval $[0, 1]$ and bouncing elastically off each other and off the "walls of the container." Recall that this apparently very simple mechanical system (with two degrees of freedom) is equivalent to the billiard flow in a right triangle with acute angle equal to arctan $(m_1/m_2)^{1/2}$.[15, 46] By this observation, every fact about billiard dynamics in right triangles corresponds to a property of this mechanical system of two elastic particles.

Finally, a remark on the references. When we quote a paper (or papers) in a theorem, it does not necessarily mean this particular theorem was proved there. Sometimes it means the theorem readily follows from the material contained in the quoted sources.

2. OVERVIEW OF TYPICAL ERGODICITY

The best available result on the ergodicity of billiards in polygons is a theorem of ref. 31. It says, roughly, that the billiard in a typical polygon is ergodic (with respect to the standard invariant measure). In order to formulate this theorem precisely, recall that a polygon P with a given number n of vertices and a fixed way of joining them by edges is determined by the positions of its vertices in the plane. Requiring that the length of ∂P be less than or equal to one, and that P contain the origin of $\mathbf{R}^2$, we identify the set $\mathscr{P}_n$ of these polygons with a compact subset of $\mathbf{R}^{2n}$ with a nonempty interior. In particular, $\mathscr{P}_n$ is a compact metric space.

A subset Y of a compact metric space X is a dense G_δ if Y is an intersection of a countable number of dense open sets. Dense G_δ sets are large subsets of X, in the sense of category,[40] as opposed to the measure-theoretic sense. A more precise statement of the theorem[31] says that the set $\mathcal{E}_n \subset \mathcal{P}_n$ of ergodic polygons is a dense G_δ. The theorem, actually, is even stronger. Let $\mathcal{P}$ be any compact set of polygons with a fixed number of sides, such that rational polygons with arbitrarily large angle denominators are dense in $\mathcal{P}$. Then the set $\mathcal{E} \subset \mathcal{P}$ of ergodic polygons is a dense G_δ. Intuitively, this means that a typical polygon in $\mathcal{P}$ is ergodic. Taking for $\mathcal{P}$ the set of right triangles, we obtain by the preceding remarks an important application to classical mechanics.

Corollary 1.[31] The mechanical system of two elastic point masses m_1 and m_2 on the interval is typically ergodic.

The proof in ref. 31 of the theorem referred to above uses the method of ref. 54 of approximating the billiard in an arbitrary polygon by rational polygonal billiards. The discussion of the proof is postponed until Section 4. For the moment note that many simple questions about the ergodicity of polygonal billiards remain wide open. For instance, it is not known whether the set of ergodic polygons has a positive measure [i.e., the standard (Lebesgue) measure in $\mathbf{R}^{2n}$, the parameter space for n-gons]. Moreover, there are no explicit examples of ergodic polygons.

3. RATIONAL POLYGONS AND DIRECTIONAL BILLIARD FLOWS

Let P be a rational polygon, and let $N = N(P)$ be the least common denominator of the angles $\pi m_i/n_i$ between the sides of P. By an elementary argument[54] the three-dimensional phase space Z of the billiard flow decomposes into the one-parameter family Z_θ, $0 \leqslant \theta \leqslant \pi/N$, of T'-invariant surfaces. A direct geometric construction associates with P a closed oriented surface $S = S(P)$ and a one-parameter family of flows T'_θ, $0 \leqslant \theta \leqslant 2\pi$, on S. For $0 < \theta < \pi/N$ there is a natural isomorphism $(S, T'_\theta) \cong (Z, T'|_{Z_\theta})$. The flows T'_θ are the *directional billiard flows* associated with P (see refs. 19 and 18 for more details).

The surface S is tiled by $2N$ copies of the polygon P. The topology of S is determined by one integer, $g(S) \geqslant 1$, the genus of S, which is easily computed from P.[18] For instance, if P is a simple polygon, then

$$g(S) = 1 + \frac{N}{2} \sum_{i=1}^{n} \frac{m_i - 1}{n_i} \tag{1}$$

where $\pi m_i/n_i$ are the vertex angles of P. It is immediate from Eq. (1) that S is a torus if and only if P tiles the plane under reflections. This happens only in four cases: P is the equilateral triangle, or P is the isoceles right triangle, or P is the triangle with the angles $\pi/2$, $\pi/3$, $\pi/6$, or P is a rectangle.

In all other cases $g(S) > 1$, and the flows T'_θ have multisaddle singularities at the special points of S. These singular or conical points of S are generated by the vertices of P with angles $\pi m/n$, $m > 1$. Every such vertex produces N/n singular points of S. Let $\sigma \in S$ be a singular point. The total angle at σ is $2\pi m$, and σ is a multisaddle with $2m$ prongs for every directional flow T'_θ (Fig. 2). The Riemannian metric on S inherited from P is flat everywhere except at the cone points of S, where the metric is singular. The flows T'_θ decompose the geodesic flow on S with respect to this singular flat metric. Each T'_θ preserves the Lebesgue measure on S induced by the metric. (The area of S is $2N$ times the area of P.)

These and other elementary fact about the flows T'_θ on S unfortunately do not help in answering the main question: for which θ is the flow T'_θ ergodic (with respect to the standard measure)? The only known approach to this question involves the sophisticated techniques of the Teichmuller theory of Riemann surfaces.[31] The flat singular metric on S

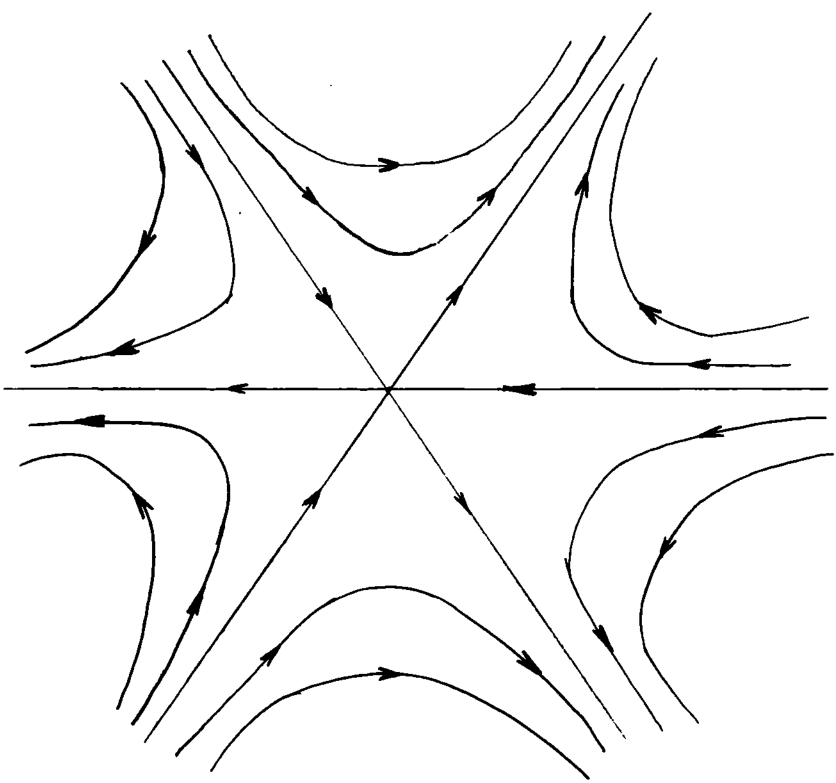

Fig. 2. Multisaddle with six prongs.

is conformally equivalent to a unique metric of curvature -1, and the flows T_θ^t correspond to a particular family of measured foliations, or quadratic differentials on S (see, e.g., ref. 45 for these notions). The date above correspond to the unit tangent bundle of the Teichmuller space $\mathscr{T}_g$ of genus g. The Teichmuller flow on $\mathscr{T}_g$ is essentially the geodesic flow with respect to the Teichmuller metric.[37] The Teichmuller flow extends to an action of the group $SL(2, \mathbf{R})$ on $\mathscr{T}_g$. The flat metric on S gets affinely distorted by this action, but remains flat. Thus, we are let to a study of the general flat metrics on S, or, equivalently, to the study of holomorphic *quadratic differentials* on a compact Riemann surface of genus g. This object is nontrivial even if $g = 1$. A detailed study of the Teichmuller flow in genus $g > 1$ leads to the following theorem, which is the main result of ref. 31.

Theorem 1.[31] Let S be a compact Riemann surface of genus $g \geqslant 1$, and let q be a holomorphic quadratic differential on S. Let T_θ^t, $0 \leqslant \theta \leqslant 2\pi$, be the family of directional flows on S, corresponding to the "rotated" quadratic differentials $e^{i\theta}q$. Then T_θ^t is uniquely ergodic for Lebesgue almost all θ.

The property of T_θ^t to be uniquely ergodic means that the Lebesgue measure on S is the only invariant measure of T_θ^t. It implies the ergodicity of T_θ^t. By the preceding material, the theorem above yields the main property of rational polygonal billiards.

Corollary 2.[31] The billiard in a rational polygon is ergodic in Lebesgue almost all directions.

The proof of Theorem 1 in ref. 31 is based on the results of ref. 33 and uses a number of facts from Teichmuller theory scattered in the literature. See ref. 1 for a more self-contained exposition.

Let P be a rational polygon, and let T_θ^t be the family of directional billiard flows. We say that a direction θ is ergodic (minimal) if the flow T_θ^t is ergodic (minimal). Recall that a flow is minimal if every infinite orbit is dense. A geometric characterization of the minimal directions (for a given polygon P) implies that the set of nonminimal directions is, at most, countable.[54] By the corollary above, the set of nonergodic directions has zero measure. It is well known that, in general, the set of nonergodic directions is larger that the set of nonminimal directions (see the references in ref. 36). For *almost integrable polygons*, the sets of minimal and ergodic directions coincide, and a directions is ergodic if and only if it is "irrational."[18, 19, 5] There are almost integrable polygons of arbitrarily high genus [see Eq. (1)], but the set of these polygons is countable.[18] A larger class of polygons for which the sets of ergodic and minimal directions coincide was found by Veech[51] (see Proposition 2 in Section 6 below). However, for a typical rational polygon the set of nonergodic directions has positive Hausdorff dimension,[39] while the set of nonminimal directions always has dimension zero, because it is countable. On the other hand, for any rational polygon the Hausdorff dimension of the set of nonergodic directions is bounded above by 1/2.[36]

Thus, in the circle of directions, the ergodic directions constitute an overwhelming majority.

4. ERGODICITY AND MIXING FOR POLYGONAL BILLIARDS

A striking application of Theorem 1 to the ergodic theory of (general) polygonal billiards is the following result.

Theorem 2.[31] For every n, there is a dense G_δ of ergodic polygons with n vertices.

Recall the terminology used in the formulation in Section 2. Here is a rough sketch of the argument. It is based on the idea of approximating a given (irrational) polygon P by a sequence P_i of rational polygons. It is intuitively clear that the billiard flow T^t of P is a limit of the sequence of flows T^t_i corresponding to P_i as $i \to \infty$. The flows T^t_i are not at all ergodic. Let M_i be a typical invariant surface for P_i, and let $D^t_i = T^t_i|_{M_i}$ be the corresponding "ergodic component." As i increases, the least common denominator $N(i)$ of the angles of P_i goes to infinity. As a consequence, the nonergodic flow T^t_i is well approximated by an ergodic component D^t_i. Thus, as $i \to \infty$, we have $D^t_i \to T^t$, yielding the assertion.

The proof of Theorem 2 in ref. 31 and the exposition in ref. 1 follow the scheme above. The approximation of an arbitrary polygon by rational polygons was first used in billiard dynamics in ref. 54. In fact, A. Katok and M. Boshernitzan pointed out to the authors of ref. 31 that the approximation technique of ref. 54 allows one to obtain Theorem 2 from Corollary 2.[28]

A similar technique of approximation is used to produce directional billiards which are *weakly mixing*.[23] Recall that the weak mixing property of ergodic theory is stronger than ergodicity and weaker than mixing (see, e.g., ref. 12 or ref. 53). A prevailing opinion in the mathematical community is that polygonal billiards are never mixing, but this has not been established. On the other hand, it seems plausible that there are weakly mixing polygons, but this also remains an open question. Results of ref. 23 provide partial evidence for this conjecture.

Denote by $\mathcal{R}_n$ the set of n-gons such that their sides are either horizontal or vertical (Fig. 3). The polygons P in $\mathcal{R}_n$ are rational, and the space $\mathcal{R}_n$, coordinatized by the lengths of the sides of P, becomes an open subset

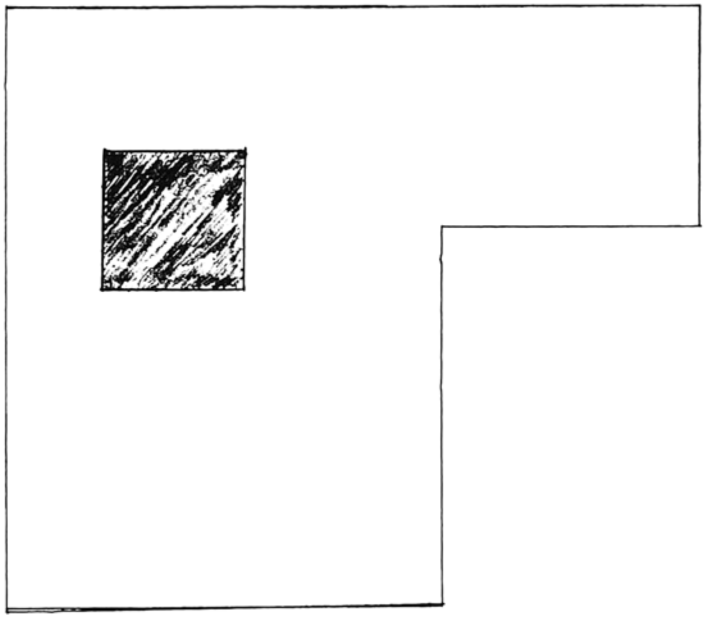

Fig. 3. Polygon whose sides are either horizontal or vertical.

of a Euclidean space. We parametrize the directions in plane by θ, $0 \leqslant \theta \leqslant 2\pi$, so that $\theta = 0$ corresponds to the positive x-axis. Then for any θ, the billiard flow $T^t_\theta(P)$ in direction θ is well defined for all polygons P in $\mathscr{R}_n$. The following theorem illustrates the results of ref. 23 on the genericity of weak mixing.

Theorem 3. (see ref. 23, Theorem 1). Let $n > 4$. Then for any direction θ, $0 < \theta < \pi/2$, the set $\mathscr{R}_{\text{mix}} \subset \mathscr{R}_n$ of polygons P such that the flow $T^t_\theta(P)$ is weakly mixing is a dense G_δ.

The proof is based on a suitable approximation argument. Here is the main idea. If the lengths of the sides of a polygon $P \in \mathscr{R}_n$ are rational, then P is an almost integrable polygon.[18] This allows one to approximate an arbitrary $P \in \mathscr{R}_n$ by a sequence $P_i \in \mathscr{R}_n$ of almost integrable rational polygons. The flows $T^t_\theta(P_i)$ have nontrivial point spectra,[18] but their size shrinks to "zero" as $i \to \infty$.

The following result supports the conjecture that polygonal billiards are not mixing.

Theorem 4.[27] For any rational polygon P and any direction θ, the directional billiard flow $T^t_\theta(P)$ is not mixing.

The proof of this theorem exploits a remarkable connection between the directional billiard flows and the *interval exchange transformations.*[12, 32] The polygonal billiard map $\phi = \phi(P)$ has special features[29] (distinguishing it from an arbitrary billiard map). If P is rational, then the phase space Φ of ϕ decomposes into a one-parameter family of ϕ-invariant subsets Φ_θ, where $0 \leqslant \theta \leqslant \pi/N$ runs through the set of directions. Under a natural isomorphism, the sets Φ_θ become intervals (we can normalize them to be $[0,1)$), and the directional billiard maps ϕ_θ become interval exchanges, ϕ_θ: $[0,1) \to [0,1)$. An interval exchange map is realized by cutting $[0,1)$ into n subintervals and rearranging them according to a permutation w on n symbols. For a fixed w, any interval exchange is determined by the n-tuple of lengths $a_1,..., a_n > 0, a_1 + \cdots + a_n = 1$, of the exchanged intervals.

By the preceding remarks, the family $T'_\theta(P)$ of directional billiard flows in a rational polygon P corresponds to a one-parameter family of interval exchanges. More generally, there is a correspondence between certain flows on surfaces and the interval exchange transformations.[2, 13] This connection has been fruitful in both directions.[3, 5–7, 9, 33, 48, 50, 27]

5. TOPOLOGY OF BILLIARD TRAJECTORIES, ORBIT CODING, AND ENTROPY

Let P be an arbitrary polygon. The billiard dynamics in P has two kinds of discontinuities. One is due to the flatness of the sides of P. It has a minor effect on the dynamics, and will not be considered in what follows. The other kind of discontinuity is caused by the vertices of P. Namely, the nearby trajectories separate, as they hit ∂P on opposite sides of a vertex. As a side effect, there are four different kinds of trajectories. The infinite trajectories are defined for all times. They never hit a vertex. The finite trajectories are only defined for a finite time interval. They begin and end at vertices (the *generalized diagonals* of P).[29] The other two kinds are the semi-infinite trajectories. They are defined either for an infinite past or for an infinite future.

In this section we study the topological properties of the trajectories of billiards in polygons. We think of them as lines in the phase space, or as piecewise linear curves in the billiard table (i.e., the configuration space). It is useful to go back and forth between these representations. It should be clear from the context which one of the two pictures we have in mind.

The standard notions of topological dynamics[53] assume that the flow (or the mapping) is continuous. Although this is not the case for billiards in polygons, the discontinuities are mild, and the basic definitions are easily modified to apply in our situation. In what follows we use the standard terminology of topological dynamics, with occasional clarifications, to account for the singularities.

Recall that a flow is *transitive* (*minimal*) if there exists a dense orbit (every infinite or semiinfinite orbit is dense). The transitivity (minimality) makes sense for the full billiard flow, as well as for directional billiard flows in a rational polygon. The following theorem is one of the earliest results on polygonal billiards.

Theorem 5.[54] The set of transitive polygons is a dense G_δ.

The proof is based on the technique of approximation by rational polygons, and uses a simple criterion for minimality of directional billiard flows.

It is natural to code billiard orbits by the sides of the polygon that they hit consecutively. We will define the necessary notions. Let $A = \{a_1,..., a_p\}$ be the set of sides of P. The *full shift space* Σ on the alphabet A consists of infinite sequences $w = \{w_i \in A: -\infty < i < \infty\}$. Analogously one defines the *one-sided shifts* $\Sigma^{\pm}$. For instance, $w \in \Sigma^+$ is given by $\{w_i \in A: 0 \leqslant i < \infty\}$. The full *shift transformation* $\sigma: \Sigma \to \Sigma$ is given by $(\sigma w)_i = w_{i+1}$, i.e., σ moves the sequence $w = \{w_i\}$ one step to the left. The one-sided shift transformations are defined analogously.

It is useful to think of elements of $\Sigma(\Sigma^{\pm})$ as infinite words in the alphabet A. It is also useful to consider the finite words $\{w_m w_{m+1} \cdots w_n;\ m < n\}$ in the alphabet A, and introduce simple operations, e.g., concatenation, on them. In particular, the meaning of the expression "a finite word v is contained in an infinite word w" should be self-explanatory.

A shift space has a natural topology: two words v and w are close if they are equal for a long time. The topology can be given by a metric (non-uniquely) which makes the shift $\sigma: \Sigma \to \Sigma$ a homeomorphism of a compact metric space. Its inverse σ^{-1} is the right shift transformation. A *subshift* $X \subset \Sigma$ is a closed subset invariant under σ. (One defines the subshifts of $\Sigma^{\pm}$ analogously.) The simplest subshifts are the *subshift of finite type* (see, e.g., refs. 32 and 42). For the convenience of the reader we describe below the subshifts of type two (Markov shifts).

Let N be $p \times p$ matrix whose entries N_{ij} are equal to one or zero (incidence matrix). The corresponding Markov shift $\Sigma_N \subset \Sigma$ consists of the infinite sequences $w = \{w_i: -\infty < i < \infty\}$ such that all length-two words $a_i a_j$ contained in w satisfy the condition $N_{ij} = 1$. Thus, if all entries of N are equal to one (zero), than $\Sigma_N = \Sigma$, the full shift ($\Sigma_N = \varnothing$, the empty subshift). Subshifts of finite type, in particular, Markov shifts, often appear in topological and smooth dynamics.[12, 32, 42, 30]

Let now $v \in \Phi$. Let $a_{i_0} a_{i_1} \cdots a_{i_n} \cdots$ be the infinite sequence of the sides of P that the *forward orbit* of v hits consecutively. Set $\beta_+(v) = \{a_{i_0} a_{i_1} \cdots a_{i_n} \cdots\} \in \Sigma^+$. This defines the *forward coding* map, $\beta_+: \Phi \to \Sigma^+$.

Reversing the time direction, we define the *backward coding* map, $\beta_+: \Phi \to \Sigma^-$, where $\beta_-(v) = \{\cdots a_{i_{-n}} \cdots a_{i_{-1}} a_{i_0}\}$ is the sequence of sides the orbit of v hit in the past. Putting the past and the future together, we define the *full coding* map, β: $\Phi \to \Sigma$.

For the sake of brevity, we will formulate the results for the full coding map, leaving the one-sided counterparts to the reader. We point out that despite its apparent simplicity and naturalness, the coding above has certain problems. One problem is obvious: the coding is not well defined on elements $v \in \Phi$ whose orbits hit the vertices of P. There is nothing we can do about it, and that is the bad part. The good part: this set is relatively small. Denote by Φ^+_{sing} (Φ^-_{sing}) the set of $v \in \Phi$ whose orbit hits a vertex of P in the future (past). Then Φ^+_{sing} (Φ^-_{sing}) is a countable union of compact analytic curves. So is the full singular set, $\Phi_{\text{sing}} = \Phi^+_{\text{sing}} \cup \Phi^-_{\text{sing}}$. The countable set $\Delta = \Phi^+_{\text{sing}} \cap \Phi^-_{\text{sing}}$ corresponds to the generalized diagonals of P. Thus the full coding β: $\Phi \backslash \Phi_{\text{sing}} \to \Sigma$ is well defined. By construction, β semiconjugates the billiard map ϕ: $\Phi \backslash \Phi_{\text{sing}} \to \Phi \backslash \Phi_{\text{sing}}$ with the shift transformation on Σ.

The other problem is that β may have "artificial discontinuities". Let v be a vertex of P with an angle π/n. A pair of parallel billiard orbits hitting ∂P arbitrarily close to v on the opposite sides remains parallel, hence close. But their codes are far apart, because they hit different sides of P. We call the vertices of P with angles π/n *removable vertices.*[20] *Unfolding* the table P around removable vertices, we associate with P a *polyhedral surface* $\tilde{P}$ that covers P. The billiard dynamics in $\tilde{P}$ projects down on the billiard in P, and the two have the same basic features. But the points in $\tilde{P}$ above the removable vertices of P are no longer singular. They are regular points, with the angle 2π. Thus $\tilde{P}$ is the *resolution of removable singularities* of P.[20] For instance, if P is one of the four types of integrable polygons, then $\tilde{P}$ is a flat torus. The billiard in P lifts to the geodesic flow on $\tilde{P}$, which is, of course, nonsingular.

To simplify the exposition, we assume in most of what follows that P has no removable vertices. By the preceding remarks, we can always reduce a polygonal billiard to this case. Thus the coding map β: $\Phi \backslash \Phi_{\text{sing}} \to \Sigma$, is well defined, and does not have artificial discontinuities. Let $\Sigma^P \subset \Sigma$ be the image of Φ under the coding map. The subshift Σ^P and the character of the mapping (onto) β: $\Phi \backslash \Phi_{\text{sing}} \to \Sigma^P$ contain a lot of information about the polygonal billiard.

The size of a subshift $X \subset \Sigma$ is measured by the growth rate of the number of words of length n contained in X as $n \to \infty$. Denote by X_n the set of words $w = (a_{i_0} \cdots a_{i_{n-1}})$ contained in X, and let $|X_n|$ be the cardinality of X_n. The exponential growth rate $h = h(X) = \limsup_{n \to \infty}[n^{-1} \log |X_n|]$ is an important characteristic of X, called the *entropy of a subshift.* It is equal to the *topological entropy* of the shift transformation on X. The entropy of a nonempty subshift of finite type is positive, e.g., the entropy of the full shift on p symbols is $\log p$.

The quantity $|\Sigma_n^P|$ has a clear dynamical meaning. Namely, $|\Sigma_n^P|$ is the number of different "histories" of the n-segment billiard orbits in P. The combinatorial length (i.e., the number of segments) of finite billiard orbits is closely related to their geometric length, but we will not use this relation. In what follows, we simply speak of the length of a finite orbit. More generally, the growth pattern of the sequence $c_n = |\Sigma_n^P|$ determines the *complexity* of the subshift Σ^P. If $h(\Sigma^P) = \limsup_{n \to \infty} \log c_n/n = 0$, then c_n grows subexponentially, and we can look for other estimates. If we can find a positive integer d and numbers $0 < a < b$ such that $an^d < c_n < bn^d$, then c_n grows as a polynomial of degree d, i.e., Σ^P has *polynomial complexity*.

Theorem 6.[29, 21, 22, 14] Let P be an arbitrary polygon. Then the entropy of the subshift Σ^P is equal to zero.

The original version of this theorem[29] assumes that P is a simply connected polygon. The work[29] analyzes the *metric entropies* with respect to the shift-invariant measures on Σ^P and proves that any such metric entropy is zero. By invoking the *variational principle*,[32, 30] ref. 29 concludes that the topological entropy of Σ^P is zero. The approach of refs. 21 and 22 is to develop a more general theory of topological entropy for a wider class of transformations with singularities: the *generalized polygon exchanges*. In particular, refs. 21 and 22 estimate the entropy of Σ^P from above by the exponential growth rate of the length of the "singular curve" of the nth iterate of the billiard transformation. A direct calculation shows that this length grows quadratically, yielding $h(\Sigma^P) = 0$.

Corollary 3. For any polygon P the subshift Σ^P is of infinite type.

A complete description of Σ^P is known only in very special cases.[26]

The coding map is not one-to-one on the set of *periodic points*. The reason is that a periodic orbit in a polygon can be moved parallel to itself, forming a *periodic strip* (see Section 6). All orbits in a periodic strip have the same, periodic, code. Is the converse true? More precisely, suppose that two forward billiard orbits $O_1 = O(v_1)$ and $O_2 = O(v_2)$ keep hitting ∂P at the same sides, yielding $\beta_+(v_1) = \beta_+(v_2)$. It is elementary to see that the *unfolded orbits* O_1^u and O_2^u are parallel.

Theorem 7.[14, 20] Let P be an arbitrary polygon. Let $v_1, v_2 \in \Phi \setminus \Phi_{\text{sing}}^+$ satisfy $\beta_+(v_1) = \beta_+(v_2)$. Then v_1, v_2 are periodic.

The version of ref. 14 assumes that P is simply connected (no obstacles). Then O_1 and O_2 are members of a band of parallel orbits,

$O = O(v)$, with the same forward code. The argument in ref. 14 shows that any such band is periodic. The version of ref. 20 makes no assumptions on P. It is then no longer true that O_1 and O_2 are contained in a band of parallel orbits, since they may be separated by an obstacle inside P. The proof in ref. 20 makes use of a global shearing property of polygonal billiards. Theorem 7 yields a number of important corollaries. We present some of them below.

Corollary 4. The coding map β: $\Phi \to \Sigma^P$ is one-to-one on the set of nonperiodic points.

Theorem 8. Let P be an arbitrary polygon. Then the metric entropy with respect to any billiard-invariant measure is zero.

We derive Theorem 8 from Theorem 7. It suffices to prove the assertion for an ergodic invariant measure μ on Φ. If the support of μ is a periodic orbit, there is nothing to prove, hence we assume that μ is supported on the set Φ_{aper} of nonperiodic points. By Theorem 7, for any $v \in \Phi_{\text{aper}}$ the forward code $\beta_+(v)$ determines v. Hence the "past" $\beta_-(v)$ is determined by the future "future" $\beta_+(v)$ for a set of full measure in Φ (in our case, all of Φ_{aper}), implying the claim.

The idea of the proof above goes back to ref. 4, which established the vanishing of the metric entropy with respect to the standard invariant measure. The argument is based on the lemma that Lebesgue almost all billiard orbits contain vertices of P in their closure.[4] The techniques above allow us to strengthen this assertion.

Theorem 9.[20] Let P be any nonintegrable polygon (i.e., P has nonremovable vertices). Then the regular points $v \in \Phi$ satisfy the following dichotomy. (i) The orbit $O(v)$ passes arbitrarily close to nonremovable vertices of P from both sides in the past and in the future. (ii) The orbit $O(v)$ is periodic.

Recall that the countable set $\Delta \subset \Phi$ consists of elements $v \in \Phi$ whose orbits are the generalized diagonals beginning and ending at the nonremovable vertices of P.

Theorem 10.[20] For any nonintegrable polygon, the set Δ is dense in Φ.

In the proof we sketch below, we assume for simplicity that all vertices of P are nonremovable. Let, for instance, $v \in \Phi_{\text{aper}}$ be a regular point. Then, by Theorem 9, the orbit $O = O(v)$ passes close to a pair of vertices $A_\pm$ in the arbitrarily remote past and future. Connecting the vertices A_-, A_+ in the unfolding, we get a sequence of generalized diagonals that converges to O.

Now we return to the question of complexity of the subshifts obtained by coding billiard orbits in a polygon P. Besides the subshift Σ^P formed by the codes of all orbits, we consider the subshifts $\Sigma^P_\theta \subset \Sigma^P$ obtained from the orbits that assume a fixed direction, $0 \leqslant \theta < 2\pi$, in their lifetime. By Theorem 6, the sequence $|\Sigma^P(n)|$ (the number of words of length n) grows subexponentially. Hence, for any $h > 0$ and sufficiently large n, we have the estimate $|\Sigma^P(n)| < e^{hn}$. There is a wide consensus that $|\Sigma^P(n)|$ grows, actually, at most polynomially. We formulate this as a conjecture.

Conjecture 1. Let P be an arbitrary polygon. Then the sequence $|\Sigma^P(n)|$ has (at most) a polynomial growth.

A more precise version of the conjecture says that for any polygon P, there is a polynomial $F_P(\,\cdot\,)$ so that

$$|\Sigma^P(n)| \leqslant F_P(n) \tag{2}$$

Condition (2) would imply (at most) polynomial growth for all kinds of dynamical quantities for polygonal billiards (see Section 6). This is certainly consistent with the estimates established for rational polygons.[20, 25] Another piece of supporting evidence for Conjecture 1 is the following theorem (due to A. Katok).

Theorem 11.[24] Let P be an arbitrary polygon. Then there is a polynomial $F_P(\,\cdot\,)$ such that for all directions $0 \leqslant \theta < 2\pi$, we have

$$|\Sigma^P_\theta(n)| \leqslant F_P(n) \tag{3}$$

6. PERIODIC ORBITS FOR BILLIARDS IN POLYGONS

In any dynamical system, the periodic orbits are important for at least two reasons. First, they are, conceptually, the simplest kind of orbit that a dynamical system can have. Second, by studying the perturbations of periodic points, we often get considerable insight into the dynamics.

Thus it is disturbing that despite the efforts of many researchers, the question of whether every polygon has a periodic orbit remains open. It is not even known if every triangle has a periodic orbit.[16] In this section we summarize the known results and the formulate a conjecture.

Let O be a periodic orbit in a polygon P, and let $n = n(O)$ be its (minimal) period. We choose a side $a \subset \partial P$ that O hits, and unfold O starting from a. Let $P_0 = P, P_1, \ldots, P_{n-1}, \ldots$ be the copies of P we obtain in the process of unfolding[19] and let O^u be the unfolded orbit. Let $x \in a$ be the point where O^u "starts" and let $x_n \in a_n \subset \partial P$ be the point it "enters" the polygon P_n. For any polygon P_i in the unfolding, there is an isometry $g_i \in Iso(\mathbf{R}^2)$

such that $P_i = g_i P_0$. By periodicity of O, we have $g_n: a \to a_n$ and $g_n(x_0) = x_n$. The analysis above and the structure of the group of isometries of $\mathbf{R}^2$ lead to the following conclusion.

Proposition 1. Let O be a periodic orbit of period n, and let the notation be as above. (i) If n is odd, then g_n is the gliding reflection about O''; it translates points of O'' by $x_0 x_n$. (ii) If n is even, then g_n is the translation by the vector $x_0 x_n$.

Proposition 1, despite its simplicity, is useful for the analysis of periodic billiards orbits. Recall that the group $Iso(\mathbf{R}^2)$ is the semidirect product of the normal subgroup of translations, which we denote simply by $\mathbf{R}^2$, and the group $O(2)$ of orthogonal 2×2 matrices. The natural homomorphism $d: Iso(\mathbf{R}^2) \to O(2)$ corresponds to taking the linear part of $g \in Iso(\mathbf{R}^2)$, and $Ker(d) = \mathbf{R}^2$. We denote by $G_P \subset Iso(\mathbf{R}^2)$ the group generated by the reflections s_i, $1 \leqslant i \leqslant p$, in the sides of P, and let $\Gamma_P = d(G_P) \subset O(2)$. The group Γ_P is generated by the matrices $\sigma_i = d(s_i) \in O(2)$, $\sigma_i^2 = id$, $1 \leqslant i \leqslant p$ (we call them reflections as well). Set $V_P \subset \mathbf{R}^2$ be the kernel of $d: G_p \to \Gamma_P$.

By Proposition 1, any periodic orbit O of odd period determines a gliding reflection $g_O \in G_P$ (nonuniquely, g_O depends on the segment of O where the unfolding starts). If O has an even period, then it determines a translation $v_O \in V_P$ (also nonuniquely). Note that the groups G_P, Γ_P, V_P are important characteristics of P. For example, P is rational (almost integrable) if and only if Γ_P is finite (G_P is discrete).[19]

Any multiple of a periodic orbit is obviously a periodic orbit. A *prime periodic orbit* is not a multiple of another one.

Corollary 5. Let O be a prime periodic orbit of period n in a polygon P. (i) If n is even, then O is contained in a band of parallel periodic orbits O_t of the same length, $l(O_t) = l(O)$. Let S be the maximal band containing O. Then S is a closed flat cylinder of length $l(O)$ and width $w = w(O) > 0$. Each of the boundary circles of S is a finite union of generalized diagonals of P. (ii) If n is odd, then the orbit O is isolated. The maximal strip S of periodic orbits parallel to O is a flat Möbius band of length $l(O)$, and O is the middle of circle of S. The boundary circle of S is a union of generalized diagonals.

The corollary makes an important distinction between the prime periodic billiard orbits of odd and even periods. The former are isolated, the latter form periodic cylinders, hence are never isolated. Periodic orbits contain a lot of geometric information about P. For example, every acute triangle has a unique periodic orbit of period 3, the Fagnano geodesic.[16, 46] The following theorem indicates, in particular, that periodic orbits of odd periods may not contribute to the statistics (of periodic orbits).

Theorem 12.[16] In any rational polygon there is (at most) a finite number of prime periodic orbits of odd periods.

An important characteristic of a dynamical system is the asymptotics of the number of periodic orbits. By the preceding results, the appropriate quantity for polygonal billiards is the number of distinct periodic cylinders of length less than l. We denote this function by $C(l)$. A separate argument shows that the set of periodic cylinders is (at most) countable.

Theorem 13.[29, 21] Let P be an arbitrary polygon. Then the function $C(l)$ grows slower than any exponential, as $l \to \infty$.

By Corollary 5, $C(l)$ is bounded above by the number of generalized diagonals of length less than l. The exponential growth rate of the number of generalized diagonals is estimated from above by the entropy of the subshift Σ^P, which is equal to zero, by Theorem 6.

Surprisingly, the theorem above is the only known upper bound on the periodic orbits in general polygons. As for a lower bound, there are none. On the other hand, periodic orbits in rational polygons have efficient bounds from below and from above. The proofs of these estimates use the theory of holomorphic quadratic differentials on Riemann surfaces.

Theorem 14.[35, 38] Let P be a rational polygon. Then there are positive constants $0 < c_1 < c_2$ such that for all sufficiently large values of l, we have

$$c_1 l^2 \leqslant C(l) \leqslant c_2 l^2 \tag{4}$$

The geometric length of a periodic orbit in equation (4) can be replaced by the number of its segments. Parallel to equation (4), there are quadratic bounds on the number of generalized diagonals.[35, 38] The quadratic upper bound on the number of generalized diagonals yields an alternative proof of Corollary 2.[5]

The lower bound in equation (4) implies, in particular, the existence of periodic orbits in rational polygons. But that can be proved in a simpler way. A billiard trajectory O is *perpendicular* if a segment of O hits an edge of ∂P at the right angle.

Theorem 15.[16, 8] Any regular perpendicular trajectory in a rational polygon is periodic.

Since all but a finite number of perpendicular trajectories are regular, the theorem above implies the existence of periodic orbits in a rational polygon. But Theorem 14 gives much more than the existence, it gives the abundance of periodic orbits for rational polygons. We also know that periodic points are dense in the phase space.[34, 10]

The proof of Theorem 15 is elementary. By recurrence and rationality, every orbit departing from a side a orthogonally to it returns to a also orthogonally. If it returns to the same point of a, it is already periodic. If it first returns to a different point, it reflects and backtracks. Hence the orbit is periodic.

The quadratic bounds (4) are not likely to hold for general (irrational) polygons. The expectation is that for general polygons there are polynomial bounds on the number of periodic cylinders.

Conjecture 2. Let P be an arbitrary polygon. Then there exist positive constants c_1, c_2 and integers $1 \leqslant n_1 \leqslant n_2$ such that for sufficiently large l we have

$$c_1 l^{n_1} < C(l) < c_2 l^{n_2} \tag{5}$$

For arbitrary rational polygons, equation (4) may well give the best possible estimates on the number of periodic cylinders. However, there are nontrivial examples when the asymptotics of $C(l)$ can be computed exactly.

Theorem 16.[51, 52] Let $n \geqslant 3$, and let P be an isoceles triangle with angles $(\pi/n, \pi/n, \pi(n-2)/n)$, or a regular n-gon. Let $\|P\|$ denote the area of P. Then there exists a positive constant $c = c(P)$, depending on the type of P, so that, as $l \to \infty$, we have

$$C(l) \cong \frac{c(P)}{\|P\|} l^2 \tag{6}$$

The explicit values of $c(P)$, computed in refs. 51 and 52, contain sophisticated number-theoretic functions. The proof of Eq. (6) involves a study of the group of *affine diffeomorphisms* of the flat surface $S = S(P)$ associated with a rational polygon. This group has a natural homomorphism into $SL(2, \mathbf{R})$. The image $\Gamma_P \subset SL(2, \mathbf{R})$ is a discrete subgroup, which is typically small. Equation (6) is related to the fact that for polygons P in Theorem 16 the groups Γ_P are large, in the sense that the quotient $SL(2, \mathbf{R})/\Gamma_P$ has a finite volume. Such subgroups are called *lattices* in $SL(2, \mathbf{R})$. One of the implications is that the directional flows T_θ^t on S satisfy a surprising dichotomy.

Proposition 2.[51] Let P be a rational polygon such that the group $\Gamma_P \subset SL(2, \mathbf{R})$ is a lattice. Then the set of directions $0 \leqslant \theta < 2\pi$ is partitioned according to the following dichotomy: Either all orbits of T_θ^t are finite (periodic or saddle connections), or the orbits of T_θ^t are all infinite.

There are many open questions suggested by the results above. For instance, can one characterize geometrically the rational polygons satisfying the dichotomy above? Can one find all rational polygons satisfying the "prime geodesic theorem" [i.e., Eq. (6)]? There are speculations that the (order of magnitude of the) error term in Eq. (6) is relevant for the quantum chaos.

We return to the periodic orbits in general (irrational) polygons. Computer simulations have suggested that the perpendicular orbits are likely to be periodic.[43] Recently this observation has been partially confirmed mathematically. We will elaborate.

The arclength along a side a of P induces a (Lebesgue) measure on the set of perpendicular trajectories departing from a. Normalizing, the measure, we can speak of the probability for a perpendicular trajectory to be periodic.

A polygon P is a *generalized parallelogram* if there are two straight lines l and m such that each side of P is parallel to either l of m. Note that a generalized parallelogram is rational if the angle between l and m is rational.

Theorem 17.[11,24] Let P be any right triangle or any generalized parallelogram. Then a perpendicular orbit in P is periodic with probability one.

The theorem yields the existence of periodic orbits for a large class of irrational polygons. The proof in ref. 24 is based on an idea of ref. 11, where the result is proved for right triangles. Using the equivalence of the billiard dynamics in right triangles and the system of two elastic masses, we obtain an application to classical mechanics.

Corollary 6. Let m_1 and m_2 be arbitrary elastic masses in an interval. Let at $t=0$ one of the masses, say m_1, be hitting the wall, while the other mass, m_2, is at rest anywhere in the interval. Then the corresponding trajectory is periodic with probability one.

NOTE ADDED IN PROOF

The long standing "illumination problem" for polygons has been solved, in the negative, by G. Tokarsky [56]. He constructs polygons, P, and pairs of points, A_0, A_1, in P, such that the billiard ball starting at A_0 will never reach A_1.

ACKNOWLEDGMENT

This work was partially supported by NSF grants DMS-9013220 and DMS-9400295.

REFERENCES

1. P. Arnoux, Ergodicité générique des billards polygonaux, Seminaire Bourbaki (1987/88), exp. 696, pp. 203–221.

2. P. Arnoux, Echanges d'intervalles et flots sur les surfaces, in *Théorie Ergodique* (Enseignement Mathématique, Geneva, 1981), pp. 5–38.
3. P. Arnoux, D. Ornstein, and B. Weiss, Cutting and stacking, interval exchanges, and geometric models, *Isr. J. Math.* **50**:160–168 (1985).
4. C. Boldrighini, M. Keane, and F. Marchetti, Billiard in polygons, *Ann. Prob.* **6**:532–540. (1978).
5. M. Boshernitzan, A condition for minimal exchange maps to be uniquely ergodic, *Duke Math. J.* **52**:725–752 (1985).
6. M. Boshernitzan, A unique ergodicity of minimal symbolic flows with linear block growth, *J. Analyse Math.* **44**:77–96 (1985).
7. M. Boshernitzan, Rank two interval exchange transformations, *Ergod. Theory Dynam. Syst.* **8**:379–394 (1988).
8. M. Boshernitzan, Billiards and rational periodic directions in polygons, *Am. Math. Monthly* **99**:522–529 (1992).
9. M. Boshernitzan and I. Cornfeld, Interval translation mappings, preprint (1994).
10. M. Boshernitzan, G. Galperin, T. Kruger, and S. Troubetzkoy, Some remarks on periodic billiard orbits in rational polygons, preprint (1994).
11. B. Cipra, R. Hanson, and A. Kolan, Periodic trajectories in right triangle billiards, preprint, St. Olaf College (1994).
12. I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory* (Springer-Verlag, Berlin, 1982).
13. A. Fathi, F. Laudenbach, and V. Poenaru, Travaux de Thurston sur les surfaces, *Asterisque* **66/67** (1979).
14. G. Galperin, T. Krueger, and S. Troubetzkoy, Local instability of orbits in polygonal and polyhedral billiards, *Commun. Math. Phys.* **169** (1995), 463–473.
15. G. Galperin, N. Chernov, and A. Zemlyakov, *Mathematics of Billiards* (Cambridge University Press, Cambridge, in press.
16. G. Galperin, A. M. Stepin, and Ya. Vorobetz, Periodic billiard orbits in polygons; generating mechanisms, *Russ. Math. Surv.* **47**:5–80 (1992).
17. G. Galperin and D. Zvonkine, Periodic billiard trajectories in right triangles and right-angled tetrahedra, preprint (1994).
18. E. Gutkin, Billiard flows on almost integrable polyhedral surfaces, *Ergod. Theory Dynam. Syst.* **4**:569–584 (1984).
19. E. Gutkin, Billiards in polygons, *Physica* **19D**:311–333 (1986).
20. E. Gutkin, Orbit coding and topology of trajectories for polygonal billiards, preprint (1995).
21. E. Gutkin and H. Haydn, Topological entropy of generalized polygon exchanges, *Bull. AMS* 32:50–57 (1995).
22. E. Gutkin and H. Haydn, Topological entropy of polygon exchange transformations and polygonal billiards, preprint, USC (1995).
23. E. Gutkin and A. Katok, Weakly mixing billiards, in *Holomorphic Dynamics*, Lecture Notes in Mathematics 1345, (Springer-Verlag, Berlin, 1989), pp. 163–176.
24. E. Gutkin and S. Troubetzkoy, Directional flows and strong recurrence for polygonal billiards, preprint (1995).
25. P. Hubert, Complexité de suites definies par des trajectories de billard, preprint (1994).
26. P. Hubert and G. Rauzy, Dynamique symbolique du billard dans le triangle equilateral, preprint (1994).
27. A. Katok, Interval exchange transformations and some special flows are not mixing, *Isr. J. Math.* **35**:301–310 (1980).

28. A. Katok, Ergodicity of generic irrational billiards, in *Abstracts of Workshop on Two-Manifold and Geometry* (MSRI, 1986), Vol. 1, p. 13.
29. A. Katok, The growth rate for the number of singular and periodic orbits for a polygonal billard, *Commun. Math. Phys.* **111**:151–160 (1987).
30. A. Katok and B. Hasselblatt, *Introduction of the Modern Theory of Dynamical Systems* (Cambridge University Press, Cambridge, 1995).
31. S. Kerckhoff, H. Masur, and J. Smillie, Ergodicity of billiard flows and quadratic differentials, *Ann. Math.* **124**:293–311 (1986).
32. R. Mañe, *Ergodic Theory and Differentiable Dynamics* (Springer-Verlag, New York, 1987).
33. H. Masur, Interval exchange transformations and measured foliations, *Ann. Math.* **115**:168–200 (1982).
34. H. Masur, Closed trajectories for quadratic differentials with an application to billiards, *Duke Math. J.* **53**:307–314 (1986).
35. H. Masur, The growth rate for trajectories of a quadratic differential, *Ergod. Theory Dynam. Syst.* **10**:151–176 (1990).
36. H. Masur, Hausdorff dimension of the set of nonergodic foliations of a quadratic differential, *Duke Math. J.* **66**:387–442 (1992).
37. H. Masur, The Teichmuller flow is Hamiltonian, *Proc. AMS*, in press.
38. H. Masur, Lower bounds for the number of saddle connections, and closed trajectories of a quadratic differential, in *Holomorphic Functions and Moduli I*, (MSRI, 1988), pp. 215–228.
39. H. Masur and J. Smillie, Hausdorff dimension of sets of nonergodic measured foliations, *Ann. Math.* **134**:455–543 (1991).
40. J. C. Oxtoby, *Measure and Category; A Survey of the Analogies Between Topological and Measure Spaces* (Springer-Verlag, New York, 1971).
41. *Quantum Chaos* (Italian Physical Society Proceedings, Course CXIX; North-Holland, Amsterdam, 1993).
42. C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, Boca Raton, Florida, 1995).
43. Th. W. Ruijgrok, Periodic orbits in triangular billiards, *Acta Phys. Polon. B* **22**:955–967 (1991).
44. P. Sarnak, Arithmetic quantum chaos, *Isr. Math. Conf. Proc.* **8**:183–236 (1995).
45. K. Strebel, *Quadratic Differentials* (Springer-Verlag, New York, 1984).
46. S. Tabachnikov, *Billiards* (Société Mathématique de France, 1995).
47. M. Troyanov, Les surfaces euclidiennes a singularités coniques, *l'Enseign. Math.* **32**:79–94 (1986).
48. W. A. Veech, Gauss measures for transformations on the space of interval exchange maps, *Ann. Math.* **155**:201–242 (1982).
49. W. A. Veech, The Teichmuller geodesic flow, *Ann. Math.* **124**:441-530 (1986).
50. W. A. Veech, Boshernitzan's criterion for unique ergodicity of an interval exchange transformation, *Ergod. Theory Dynam. Syst.* **7**:149–153 (1987).
51. W. A. Veech, Teichmuller curves in modular spaces, Eisenstein series, and an application to triangular billiards, *Inv. Math.* **97**:553–583 (1989).
52. W. A. Veech, The billiard in a regular polygon, *Geom. Funct. Anal.* **2**:341–379 (1992).
53. P. Walters, *An Introduction to Ergodic Theory* (Springer-Verlag, New York, 1982).
54. A. Zemlyakov and A. Katok, Topological transitivity of billiards in polygons, *Math. Notes* **18**:760–764 (1975).
55. K. Zyczkowski, Classical and quantum billiards: Integrable, nonintegrable, and pseudo-integrable, *Acta Phys. Polon. B* **23**:245–270 (1992).
56. G. W. Tokarsky, Polygonal rooms not illuminated from every point, *Amer. Math. Monthly* **102**:867–879 (1995).