WHAT IS...

a Billiard?

Yakov Sinai

Billiards are dynamical systems. In the simplest case, a "billiard table" is a compact domain $Q \subset R^d$ with a piecewise smooth boundary. For a large part of the theory the class of smoothness plays no role. The reader is invited to think about components of the boundary ∂Q as subsets of C^∞-submanifolds of codimension 1. The phase space M of the billiard is the unit tangent bundle of Q with the natural identification at the boundary

$$(1) \qquad v' = v - 2(v, n(q))n(q), \quad q \in \partial Q$$

where $n(q)$ is the inward-pointing unit normal vector at q. The reflection law (1) is not used at intersections of the several components of the boundary, where it has no meaning.

The dynamics $\{S^t\}$ of a billiard is the uniform motion with unit speed inside Q of a point representing the billiard ball and with the reflections off the boundary given by (1), $-\infty < t < \infty$. The group $\{S^t\}$ preserves the Liouville measure $dq d\omega_q(v)$ where ω_q is the uniform measure on the unit sphere of tangent vectors to Q with given q.

Mechanical systems with elastic collisions often give rise to billiards. For example, a system of one-dimensional point particles with arbitrary masses moving freely between elastic collisions can be described as a billiard system inside a simplex whose dimension equals the number of particles. The system of N hard balls of radius ρ in a volume V is reduced to the billiard in the domain $\underbrace{V \times V \times \cdots \times V}_{N \text{ times}} \setminus C$ where C is the union of the cylinders

$$C_{ij} = \left\{ q^{(1)}, \ldots, q^{(N)}, |q^{(i)} - q^{(j)}|^2 \leq (2\rho)^2 \right\}.$$

All properties of a billiard system are determined entirely by the geometric properties of ∂Q. In particular, the curvature of the configuration space Q is concentrated at the boundary. In the two-dimensional setting billiards are intermediate between geodesic flows and flows generated by quadratic differentials where the curvature is concentrated at isolated points.

The most thoroughly studied billiards are the two-dimensional billiards. Some are integrable, meaning that the phase space M, minus some submanifolds of smaller dimension, can be decomposed into two-dimensional invariant tori, and the dynamics on each torus is described by quasi-periodic functions. Examples of such billiards include: (1) billiards inside parallelograms, (2) billiards inside equilateral triangles, and (3) billiards inside ellipses. According to a popular conjecture, the set of integrable billiards can be fully described and is not much wider than this list.

Billiards in general strictly convex smooth domains have some properties of integrable billiards. A curve $\gamma \subset Q$ is called a caustic if any tangent ray to γ after reflection remains tangent to γ. Caustics play an important role, because some semiclassical approximations of eigenfunctions of Laplacians are described in terms of caustics. V. F. Lazutkin has shown that the set of tangent vectors to all caustics is a set of positive measure in the phase space, accumulating near the boundary. In spirit this result belongs to KAM-theory. J. Mather proved that if the curvature of ∂Q is zero at one or several points and negative otherwise, the billiard has no caustics (remember the orientation of $n(q)$).

If Q is a polygon whose angles are rational multiples of π, then the velocity along each trajectory of the billiard can take finitely many values. Fixing these values we get a vector field on a two-dimensional surface whose trajectories can be represented as canonical foliations generated by quadratic differentials on Riemann surfaces. H. Masur and W. Veech have shown that typically these fields are strictly ergodic; i.e., they admit a unique invariant measure that naturally is ergodic. The same statement holds true for the so-called interval exchange transformations (IET). These are maps of $[0, 1]$ that arise if one cuts $[0, 1]$ into several subintervals $\Delta_1, \Delta_2, \ldots, \Delta_k$ and puts them in a different order according to some permutation.

Yakov Sinai is professor of mathematics at Princeton University. His email address is sinai@math.princeton.edu.
The author thanks E. Lieb for useful remarks.

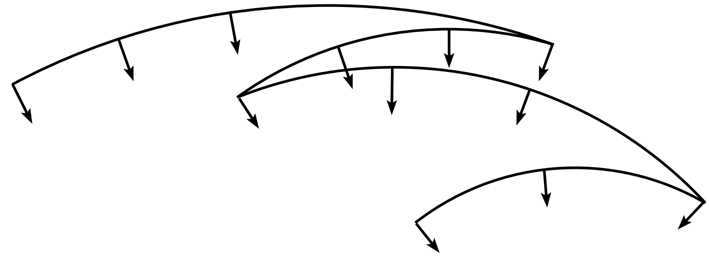
IETs are closely connected with billiards in polygons. The theory of such billiards is now an actively studied topic in the theory of dynamical systems. Almost nothing is known if the angles of a polygon are incommensurate with π. Even the theory of billiards in triangles with two irrational angles awaits development.

Much can be said if the curvature of ∂Q is strictly positive on some components of the boundary and is zero on the others. Such billiards are called hyperbolic. This condition links the theory of hyperbolic billiards with geodesic flows on manifolds of negative curvature, Anosov flows, and the general theory of hyperbolic dynamical systems. A similar definition can be easily given in the multidimensional case. Probably the first person who noticed the analogy between systems with elastic collisions and related billiards and geodesic flows on manifolds of negative curvature was the Soviet physicist N. S. Krylov. In the two-dimensional case the simplest examples of hyperbolic billiards are squares from which one or several strictly convex scatterers are cut out.

L. Bunimovich showed that the billiard inside a "stadium" is in a natural sense also a hyperbolic billiard. A stadium is a domain bounded by two semicircles and two parallel straight segments. Later, Bunimovich and Donnay extended this result to a wider class of domains in which semicircles can be replaced by general strictly convex curves and the straight segments are allowed to be nonparallel. The motion of a billiard ball on a table that is the complement of a random or periodic configuration of strictly convex scatterers is called a Lorentz gas and is one of the most popular models in nonequilibrium statistical mechanics.

Bunimovich stadia and similar billiards are popular models in the theory of quantum chaos, which studies the connections between eigenfunctions of Laplacians and ergodic properties of underlying classical dynamical systems. One of the reasons is their simplicity and amenability to numerical methods.

In the hyperbolic theory of dynamical systems a stable (unstable) manifold of a point $x \in M$ is a local submanifold $\gamma^{(s)}(x)$ ($\gamma^{(u)}(x)$) such that $\mathrm{dist}(S^t y, S^t x) \le C(x)\exp\{-\lambda|t|\}$, $t > 0$ ($t < 0$), for all $y \in \gamma^{(s)}(x)$ ($y \in \gamma^{(u)}(x)$), where $C(x)$ and λ are positive numbers. In the case of geodesic flows on manifolds of negative curvature, stable and unstable manifolds are horocycles ($d = 2$) and horospheres ($d > 2$). In the case of hyperbolic billiards, almost every point also has a stable (*sm*) and unstable (*um*) manifold. This statement is a particular case of the general Hadamard-Perron theorem and is a relatively simple part of the theory. The new feature compared to the smooth situation is the appearance of cusp-type singularities on these manifolds, which are created by trajectories that were tangent to the boundary at some time in the past. A typical form of *sm* is given in the accompanying figure.



A typical form of a stable manifold.

This existence of *sm* and *um* is a manifestation of the intrinsic instability of the dynamics. Therefore, hyperbolic billiards are among the most popular models of deterministic chaos.

The main problem related to hyperbolic billiards is the problem of their ergodicity. There is a general argument due to E. Hopf that gives the ergodicity if *sm* and *um* have a property called "local transitivity", meaning that for any two close points $x, y \in M$ one can construct a continuous path from x to y which consists of finitely many components such that each component belongs either to *sm* or to *um*. For smooth systems where the sum of dimensions of *sm* and *um* is $2d - 2$, local transitivity follows directly from their general properties. It is not so for billiards, because smooth components of *sm* or *um* can be arbitrarily small. This difficulty can be overcome with the help of the Fundamental Theorem for hyperbolic billiards, which has several versions. The first one says that in an arbitrary small neighborhood U of a typical point and an $sm \subset U$, the probability (in a natural sense) of points in *sm* for which the size of *um* is not smaller than the size of the initial *sm* is greater than some constant. This property is enough to carry out Hopf's argument.

In the second version, based on the so-called Chernov-Sinai Ansatz, it is shown that local transitivity holds because there are no submanifolds of codimension 1 that separate different subsets with the property of local transitivity; therefore, this subset is unique.

Further Reading

[1] *Dynamical Systems, Ergodic Theory and Applications* (Y. G. Sinai, ed.), Encyclopedia of Mathematical Sciences, vol. 100, Springer-Verlag, Berlin, 2000.

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

[3] *Selected Works of Eberhard Hopf with Commentaries* (Cathleen S. Morawetz, James B. Serrin, and Yakov G. Sinai, eds.), Collected Works, vol. 17, Amer. Math. Soc., Providence, RI, 2002.