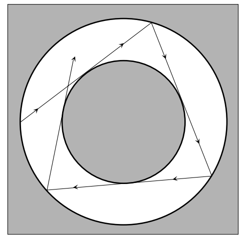
PANORAMAS ET SYNTHÈSES 1

BILLIARDS

Serge Tabachnikov

Société Mathématique de France 1995



BILLIARDS

Serge Tabachnikov

This survey of the theory of mathematical billiards reflects the present state in this very active research area. The author attempted to make the exposition as geometrical as possible. Smooth convex billiards, billiards in polygons and polyhedra, and chaotic dispersing billiards are treated in separate chapters. Another chapter concerns a lesser-known topic of dual (or outer) billiards. The survey is intended for non-experts and provides a general mathematical background for the study of billiards.

Ce survol de la théorie des billards mathématiques reflète l'état de ce champ très actif de recherches. L'auteur a essayé de rendre l'exposition aussi géométrique que possible. Les billards convexes, lisses, les billards dans les polygones ou polyèdres et les billards à dispersion chaotique sont traités dans différents chapitres. Un autre chapitre est consacré au problème moins connu des billards duaux (ou extérieurs). Le texte est destiné à des non-spécialistes et fournit l'arrière-plan mathématique nécessaire à l'étude des billards.

AMS subject classification : 58 F, 70 H.

CONTENTS

1. INTRODUCTION

Mathematical billiards is a rich and beautiful subject. It is very extensive as well. The choice of material for this survey reflects the taste of the author, who has attempted to make the exposition as geometrical as possible.

The survey consists of five chapters :

- the first provides some general background ;
- the second concerns convex smooth billiards (elliptic case) ;
- the third deals with billiards in polygons and polyhedra (parabolic case) ; the fourth discusses the lesser-known topic of dual billiards, which are of particular interest to the author ; and
- the fifth is a very brief treatment of chaotic billiards (hyperbolic case).

Each chapter has a brief introduction of its own and is subdivided into sections ; it goes without saying that «Lemma 1.2.3» means «Lemma 3 from Section 2, Chapter 1».

The author is grateful to the mathematicians he had the opportunity to discuss billiards with and learn from : V. Arnold, M. Audin, M. Berger, M. Bialy, Ph. Boyland, N. Chernov, D. Fuchs, G. Galperin, E. Ghys, A. Givental, M. Gromov, E. Gutkin, P. Iglesias, A. Katok, R. de Llave, J. Moser, Ya. Pesin, L. Polterovich, Ya. Sinai, J. Smillie, S. Troubetzkoy, A. Veselov, M. Wojtkowski. They taught him far more than these notes show.

Special gratitude goes to I. Monroe for his help with the numerical study of dual billiards, to J. Duncan for patiently reading the text and giving stylistic advice and to E. Gutkin, G. Galperin and the referees whose suggestions helped to improved the exposition. Last but not least, it is a pleasure to acknowledge the partial support of an ASTA grant 94–B–25.

1. GENERAL THEORY AND MATHEMATICAL BACKGROUND

This chapter concerns the definition of the billiard transformation and the billiard flow. It also provides a necessary mathematical background for the rest of the survey.

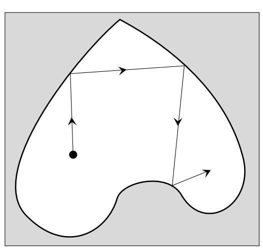
Section 2 introduces the area form invariant under the billiard transformation in the two-dimensional case, and the variational approach to the billiard problem.

Section 3, in the spirit of the geometrical optics, deals with the billiard transformation of the space of rays in the plane.

Section 5 is a very brief introduction to symplectic geometry, which we use in the next section to understand the results of Sections 2 and 3 from a more general viewpoint and to generalize them to the higher-dimensional case.

Section 7 concerns discontinuities of the billiard transformation ; we show that the transformation and the billiard flow for all times are defined almost everywhere in the sense of measure. Section 10 introduces the hierarchy of stochastic properties such as ergodicity, minimality, topological transitivity, mixing, etc. The contents of other section is self-explanatory.

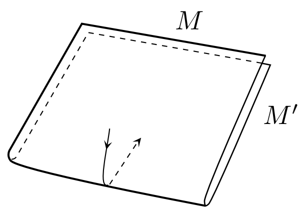
1.1 Mathematical Billiards. Phase and Configuration Spaces



A billiard table is a Riemannian manifold M with a piecewise smooth boundary. The billiard dynamical system in M is generated by the free motion of a mass-point (called a billiard ball) subject to the elastic reflection in the boundary. This means that the point moves along a geodesic line in M with a constant (say, unit) speed until it hits the boundary. At a smooth boundary point the billiard ball reflects so that the tangential component of its velocity remains the same, while the normal component changes its sign. In dimension two this collision is described by a well known law of geometrical optics : the angle of incidence equals the angle of reflection. Thus the theory of billiards and the theory of geometrical optics have many features in common. If the billiard ball hits a corner, its further motion is not defined (there are some exceptions to this to be discussed later).

The time-t billiard transformation acts on unit tangent vectors to M, more precisely, on those pairs (x, v) with $x \in M$, $v \in T_xM$ whose trajectories undergo finitely many reflections in the boundary and avoid corners on the time interval $[0, t]$. The unit tangent bundle to M is the phase space of the billiard, and the manifold M is its configuration space.

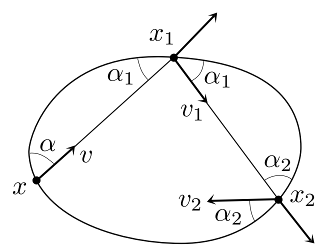
Billiards are the geodesic flows on Riemannian manifolds with boundaries. They can also be treated as a limit case of the geodesic flows on boundaryless manifolds, at least heuristically. Let M be a smooth plane billiard table. Consider its «thickening», i.e. an infinitely thin three dimensional body whose boundary N is obtained by pasting two copies of M along their boundaries and smoothening the edge. Then a billiard trajectory in M can be viewed as a geodesic line on the boundary of N, that goes from one copy of M to another each time the billiard ball bounces off the boundary. This construction is due to G. Birkhoff [Bi1].



1.2 Invariant Measure and Generating Function for Plane Billiards. Variational Formulation

So far billiards were defined as a continuous time dynamical system. One can reduce the dimension by one and replace continuous time by discrete time, i.e. replace a flow by a mapping.

Let M be a bounded plane billiard table. Consider the manifold V of unit tangent vectors (x, v) with the inward direction v and the footpoint x on the boundary ∂M. If the boundary consists of one component then V is an annulus $S^1 \times I$. We now define the billiard transformation T of V. A vector (x, v) moves along the straight line through x in the direction of v to the next point of its intersection x_1 with ∂M, and then v reflects in ∂M to a new vector $v_1 : T(x, v) = (x_1, v_1)$.



A very remarkable property of the billiard transformation T is the existence of an invariant area form. Parametrize ∂M by the length parameter t and let α be the angle between v and the direction of the boundary at the point $x(t)$. Use (t, α) as the coordinates in V, $\alpha \in [0, \pi]$.

Lemma 1.2.1. — *The area form* $\sin\alpha \, \mathrm{d}\alpha \wedge \mathrm{d}t$ *is* T*-invariant.*

Proof. — Let $T(t, \alpha) = (t_1, \alpha_1)$. One wants to prove that

$$\sin\alpha_1 \, \mathrm{d}\alpha_1 \wedge \mathrm{d}t_1 = \sin\alpha \, \mathrm{d}\alpha \wedge \mathrm{d}t.$$

Let $H(t, t_1)$ be the distance between the points $x(t)$ and $x(t_1)$. It readily follows from elementary geometry that

$$\frac{\partial H(t, t_1)}{\partial t} = -\cos\alpha \quad \text{and} \quad \frac{\partial H(t, t_1)}{\partial t_1} = \cos\alpha_1.$$

Hence

$$\cos\alpha_1 \,\mathrm{d}t_1 - \cos\alpha \,\mathrm{d}t = \frac{\partial H(t,t_1)}{\partial t_1}\,\mathrm{d}t_1 + \frac{\partial H(t,t_1)}{\partial t}\,\mathrm{d}t = \mathrm{d}H,$$

and taking differentials,

$$\sin\alpha_1 \; d\alpha_1 \wedge \mathrm{d}t_1 - \sin\alpha \,\mathrm{d}\alpha \wedge \mathrm{d}t = 0. \quad \square$$

Consider three consecutive points :

$$(t_1, \alpha_1) = T(t, \alpha), \quad (t_2, \alpha_2) = T(t_1, \alpha_1).$$

It follows from the previous proof that

$$\frac{\partial H(t,t_1)}{\partial t_1} = \cos\alpha_1, \; \frac{\partial H(t_1,t_2)}{\partial t_1} = -\cos\alpha_1.$$

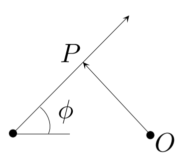
Hence

$$\frac{\partial H(t,t_1)}{\partial t_1} + \frac{\partial H(t_1,t_2)}{\partial t_1} = 0.$$

This formula has the following interpretation. Suppose that one wants to start the billiard ball at the point x so that after one reflection in the boundary at some point x_1 it arrives to the given point x_2. How does one find the unknown point x_1 ? Answer : this point is a critical point of the functional $\mathrm{dist}(x,x_1)+\mathrm{dist}(x_1,x_2)$. This variational principle plays an important role in the study of billiards.

1.3 Billiard Transformation of the Space of Rays in the Plane

It would be more in the spirit of geometrical optics to deal with oriented lines (or rays). Such an approach to billiards is possible and indeed fruitful.



Suppose that a plane billiard table M is convex (this assumption is not really necessary ; what follows will hold true locally for a generic M). Let U be the set of oriented lines in the plane that intersect M. To parametrize the set of rays choose the origin O inside M. Given a ray ℓ, drop the perpendicular OP onto it. Fix a direction in the plane and let ℓ make the angle of ϕ with it. Let $p = \pm|OP|$ depending on the orientation of the frame $(\vec{\ell}, \overrightarrow{OP})$. Then (p, ϕ) are the coordinates in the set of oriented lines in the plane. The set U is given by the inequality $|p| \leq f(\phi)$, where the function $f(\phi)$ depends on the shape of M (it is called the *support function*). It follows that U is diffeomorphic to an annulus.

Define the billiard transformation T' of U : the ray, that contains a segment of the trajectory of the billiard ball, oriented by the direction of its motion, is sent to the ray, that contains the next segment of this trajectory after the reflection in the boundary.

The set of lines in the plane is an object of study in integral geometry. It is known that there exists a unique, up to a constant factor, measure on the set of lines invariant under the motions of the plane. In our notation this is given by the 2-form $\mathrm{d}p \wedge \mathrm{d}\phi$ (see [Sa], and also [Ig] on the relation between symplectic and integral geometry).

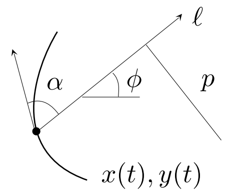
Identify the manifold V, introduced in the previous section, with U : the ray through x in the direction of v corresponds to a point $(x, v) \in V$. Thus one identifies the transformations T and T'. Compare the two 2-forms $\sin\alpha\,\mathrm{d}\alpha \wedge \mathrm{d}t$ and $\mathrm{d}p \wedge \mathrm{d}\phi$.

Lemma 1.3.1. — *These forms are equal.*

Proof. — An equation of the line, whose coordinates in U are (p, ϕ), is

$$y\cos\phi - x\sin\phi = p.$$

Differentiate : $\cos\phi\,\mathrm{d}y - \sin\phi\,\mathrm{d}x - (y\sin\phi + x\cos\phi)\,\mathrm{d}\phi = \mathrm{d}p$. Hence



$$\cos\phi\,\mathrm{d}y \wedge \mathrm{d}\phi - \sin\phi\,\mathrm{d}x \wedge \mathrm{d}\phi = \mathrm{d}p \wedge \mathrm{d}\phi.$$

The angle made by the direction of ∂M at the point $x(t), y(t)$ and the fixed direction is $\alpha + \phi$. Therefore

$$\mathrm{d}y = \sin(\alpha + \phi)\,\mathrm{d}t, \qquad \mathrm{d}x = \cos(\alpha + \phi)\,\mathrm{d}t.$$

Hence

$$\big(\cos\phi\sin(\alpha+\phi) - \sin\phi\cos(\alpha+\phi)\big)\,\mathrm{d}t \wedge \mathrm{d}\phi = \sin\alpha\,\mathrm{d}t \wedge \mathrm{d}\phi = \mathrm{d}p \wedge \mathrm{d}\phi.$$

Since $\mathrm{d}(\alpha + \phi)/\mathrm{d}t = K(t)$, the curvature of the curve ∂M, one has :

$$\mathrm{d}\phi = -\,\mathrm{d}\alpha + K\,\mathrm{d}t.$$

Therefore $\mathrm{d}t \wedge \mathrm{d}\phi = \mathrm{d}\alpha \wedge \mathrm{d}t$ and, finally,

$$\sin\alpha\,\mathrm{d}\alpha \wedge \mathrm{d}t = \mathrm{d}p \wedge \mathrm{d}\phi. \quad \square$$

It follows that the billiard transformation T' preserves the natural area form of the space of rays in the plane.

1.4 Language of Symplectic Geometry

To put the observations, made so far about plane billiards, into a proper context, and to generalize them to higher dimensions, one needs some basic concepts of symplectic geometry (see the excellent surveys [A-G], [Ar1] and [Gro]).

Definition. — A *symplectic manifold* (M, ω) is a smooth manifold M with a closed nondegenerate differential 2-form ω, called the *symplectic structure*.

Since the 2-form is nondegenerate, the dimension of a symplectic manifold is even. A symplectic manifold has a canonical volume form ω^n, where $2n = \dim M$.

Example

A $2n$–dimensional linear space with coordinates $x_1, ..., x_n, y_1, ..., y_n$ has a linear symplectic structure $\mathrm{d}x \wedge \mathrm{d}y = \mathrm{d}x_1 \wedge \mathrm{d}y_1 + \cdots + \mathrm{d}x_n \wedge \mathrm{d}y_n$.

Unlike Riemannian manifolds symplectic manifolds are all locally equivalent (Darboux's theorem) : there exists a local diffeomorphism that carries one symplectic form to another. Hence the previous example provides a local normal form of a symplectic manifold. The coordinates x, y in which $\omega = \mathrm{d}x \wedge \mathrm{d}y$ are called *Darboux coordinates.*

The following example is of fundamental importance to classical mechanics, and in particular, to the theory of billiards.

Example

The cotangent bundle T^*M of a smooth manifold M has a symplectic structure. Let λ be the tautological differential 1-form on T^*M (called the Liouville form), i.e., the form whose value on a vector ξ, tangent to T^*M at a point (q, p) with $q \in M,\ p \in T^*M$, is equal to the value of the covector p on the projection of ξ to the tangent space T_qM. The natural symplectic structure on T^*M is the 2-form $\omega = \mathrm{d}\lambda$.

Identify a linear $2n$-dimensional space with the cotangent bundle of an n-dimensional space, and choose the «position» coordinates $q_1, ..., q_n$ and the dual «momentum» coordinates $p_1, ..., p_n$. In these coordinates

$$\lambda = p\,\mathrm{d}q = p_1\,\mathrm{d}q_1 + \cdots + p_n\,\mathrm{d}q_n,$$
$$\omega = \mathrm{d}p \wedge \mathrm{d}q = \mathrm{d}p_1 \wedge \mathrm{d}q_1 + \cdots + \mathrm{d}p_n \wedge \mathrm{d}q_n.$$

Definition. — An n-dimensional submanifold L of a symplectic manifold (M^{2n}, ω) is called *Lagrangian* if the restriction of ω to L vanishes.

Since ω is a nondegenerate 2-form, n is the greatest possible dimension of a submanifold on which the symplectic form vanishes.

Examples

A smooth curve in the plane (with any area form) is a Lagrangian manifold. Fibers of a cotangent bundle are Lagrangian manifolds. Given a smooth function on a manifold M the graph of its differential, considered as a section of the cotangent bundle, is a Lagrangian submanifold in T^*M.

Definition. — A diffeomorphism of symplectic manifolds that carries one symplectic structure to another is called a *symplectomorphism.*

Symplectomorphisms carry Lagrangian manifolds to Lagrangian manifolds. If $f\colon (M_1, \omega_1) \to (M_2, \omega_2)$ is a symplectomorphism then its graph is a Lagrangian submanifold of the product manifold $M_1 \times M_2$ with the symplectic structure $\omega_1 \ominus \omega_2$. The converse is also true.

Definition. — A symplectic manifold is called *exact* if its symplectic structure is the differential of a 1-form : $\omega = \mathrm{d}\lambda$.

Cotangent bundles are exact symplectic manifolds. An exact symplectic manifold cannot be compact since its symplectic volume form is exact. Let T be a symplectomorphism of an exact symplectic manifold. Then $T^*\omega = \omega$, or $\mathrm{d}(T^*\lambda - \lambda) = 0$.

Definition. — A symplectomorphism is called *exact* if the closed 1-form $(T^*\lambda - \lambda)$ is exact

$$T^*\lambda - \lambda = \mathrm{d}H$$

for a function H. The function H is called a *generating function* of the exact symplectomorphism.

To a function on a symplectic manifold a vector field naturally corresponds (the function is sometimes called a *Hamiltonian function* and the field — the *Hamiltonian vector field*). A symplectic structure, being a nondegenerate 2-form, defines an isomorphism between the tangent and the cotangent bundles. The Hamiltonian vector field (a section of the tangent bundle) corresponds under this isomorphism to the differential of the Hamiltonian function (a section of the cotangent bundle). In Darboux coordinates the Hamiltonian field of a function f is

$$-\frac{\partial f}{\partial y}\frac{\partial}{\partial x} + \frac{\partial f}{\partial x}\frac{\partial}{\partial y}\cdot$$

The flow of a Hamiltonian vector field preserves the symplectic structure. In particular, it preserves the symplectic volume (Liouville's theorem). The Hamiltonian vector field of a function f is also called its symplectic gradient and denoted by $\operatorname{sgrad} f$.

Hamiltonian vector fields form a Lie algebra under the usual commutator of vector fields. Hamiltonian functions also form a Lie algebra whose operation is called *Poisson bracket.* The Poisson bracket of two functions f and g is the derivative of one of them along the Hamiltonian vector field of the other. In Darboux coordinates

$$\{f, g\} = \frac{\partial f}{\partial x}\frac{\partial g}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial g}{\partial x}\cdot$$

The map $f \mapsto \operatorname{sgrad} f$ is a homomorphism of Lie algebras.

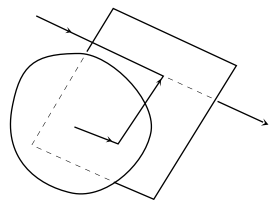
Given a hypersurface in a symplectic manifold, the restriction of the symplectic structure to it is not nondegenerate any more : it has a one-dimensional kernel in each tangent hyperplane.

Definition. — This kernel is called the *characteristic direction.* Integral lines of the field of characteristic directions are called *characteristic lines*, or simply characteristics. Characteristic lines constitute the *characteristic foliation* of a hypersurface.

If a smooth function is constant on a hypersurface then its symplectic gradient is tangent to the hypersurface along its characteristic lines.

Suppose that the set of characteristics of a hypersurface in a symplectic manifold is itself a manifold (locally it is always the case). Then this new manifold of characteristics carries a symplectic structure : its value at a pair of tangent vectors $\bar{\xi}$ and $\bar{\eta}$ equals the value of the original symplectic structure at vectors ξ and η, tangent to the hypersurface at some point, that project to $\bar{\xi}$ and $\bar{\eta}$ (the result does not depend on the choice involved).

Let M be a Riemannian manifold. Consider the hypersurface in T^*M that consists of unit covectors. The characteristics of this hypersurface are identified with nonparametrized oriented geodesic lines in M. Hence if the set of oriented geodesics of M is a manifold, it is a symplectic manifold.



Example

The set of oriented lines in a Euclidean space is a symplectic manifold. Up to the sign of the symplectic structure it is symplectomorphic to the cotangent bundle of the unit sphere in the space. The diffeomorphism of the space of rays with T^*S^{n-1} is shown in the figure (exercise for the reader : formulate it explicitly).

1.5 Symplectic Properties of Billiards

Now we are in a position to reconsider the results of Sections 1.2–1.3 from a broader point of view. To fix ideas, consider a bounded strictly convex domain M with a smooth boundary in an n-dimensional Euclidean space. This will be the billiard table (in the general case the results of this section hold locally, that is, in a neighbourhood of a generic point of the phase space). Identify the tangent and cotangent bundles using the Euclidean structure, and consider the natural symplectic structure of T^*M. Two hypersurfaces in T^*M are of importance : the one that consists of unit (co)vectors, and the one that consists of (co)vectors whose footpoints lie on the boundary ∂M. Call these hypersurfaces Y and Z, correspondingly. Their intersection W consists of unit tangent vectors with the footpoints on ∂M. This intersection is transversal.

Consider the characteristic foliations of the hypersurfaces Y and Z. We already know that the characteristics of Y are oriented lines in the space, that intersect M. Let U be the manifold of these lines; it has the induced symplectic structure, introduced in the previous section. U is diffeomorphic to the unit disc subbundle of the cotangent bundle of the unit sphere in the space.

To describe the characteristics of Z, consider the map $Z \to T(\partial M)$ that projects a tangent vector in the ambient space onto the tangent hyperplane to ∂M.

Lemma 1.5.1. — *The characteristics of Z are the fibers of this projection.*

Proof.—Introduce the usual coordinates $q_1, ..., q_n, p_1, ..., p_n$ in the cotangent bundle of the space; $\omega = \mathrm{d}p \wedge \mathrm{d}q$. Let $f(q) = 0$ be a local equation of ∂M. Then the fibers of the projection $Z \to T(\partial M)$ are generated by the normal vectors to ∂M, i.e., by the vectors $\dfrac{\partial f(q)}{\partial q}\dfrac{\partial}{\partial p}$. The inner product of this vector with ω equals $\mathrm{d}f$ — a 1-form that vanishes on tangent hyperplanes to ∂M. $\square$