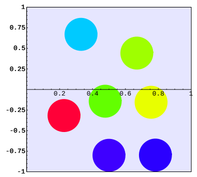
CHAOTIC BILLIARDS — Math118, O. Knill

ABSTRACT. Billiards in tables with negative curvature as well as billiards like the Stadium are chaotic: The Lyapunov exponent is positive. They are actually ergodic: every invariant set of positive measure will have either area 0 or area 1.

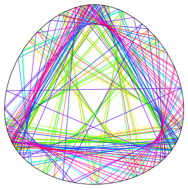
POINCARES RECURRENCE THEOREM. Area preservation allows to make a statement about recurrence of area-preserving map defined on a T invariant subset in the plane. For example, X could be the annulus $R/Z \times [-1,1]$ and T could be a billiard map.

> For every set Y of positive area $|Y|$, there exists n such that $T^n(Y) \cap Y$ has positive area.

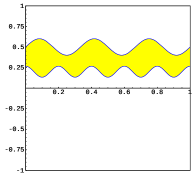
PROOF OF POINCARES THEOREM. Assume there exists a set Y of positive area $m(Y)$ such that $Y_i = T^i(Y)$ satisfies $m(Y_i \cap Y) = 0$ for all $i > 0$. Because $m(Y_i) = m(Y) > 0$ and the total space has finite area, there must exist $0 < i < j$ such that $m(Y_i \cap Y_j) > 0$. (This is a variant of the pigeon hole principle. If you have a cage with finite room and each pigeon needs the same amount of space, only a finite number of pigeons fit). But $m(T^{-i}(Y_i \cap Y_j)) = m(Y \cap Y_{j-i}) > 0$ contradicts that Y and Y_k are disjoint.



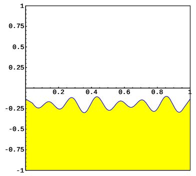
CONSEQUENCE FOR BILLIARDS. Does this mean that if you start shooting from a certain point in a certain direction, there will be times, when the orbit will come back to a similar spot on the table with a similar angle? Not necessarily. For example, if you are on the stable manifold of an unstable periodic point, then the orbit will converge to that periodic orbit. The Poincaré statement is a statement about sets. It assures for example, that if you start shooting from a certain interval on the table in a certain interval of directions, you will come back to that range of initial conditions **with probability 1**.



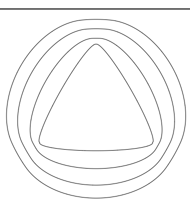
ERGODICITY. Less obvious is the question, whether a given set ever reaches an other set. If all "measurable" invariant subset of the annulus have either area 1 or 0, then the map is called **ergodic**. Measurable is a technical term which assures that the area $\int \int_A 1 \, dxdy$ is defined. Any set which can be defined by a (possibly infinitely) sequence of intersections or unions is measurable.



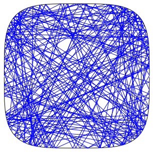
INVARIANT CURVES PREVENT ERGODICITY. If a billiard has an invariant curve which is the graph of a function $\{y = f(x)\}$, then if (x_0, y_0) is below the graph, the entire orbit (x_n, y_n) stays below the graph for all times. The billiard can not be ergodic.



STRING CONSTRUCTION. It had been known since a long times, that if one starts with a convex curve, winds a closed string around it and drags the string around the curve which keeping the string tight, we obtain a table, which has the original curve as a caustic. The picture shows some tables which have a triangle as a caustic. These tables are not ergodic.



GLANCING BILLIARDS. An orbit (x_j, y_j) of a billiard table for which y_j comes arbitrarily close to -1 and arbitrarily close to 1 is called a **glancing billiard orbit**.



> THEOREM. (Birkhoff) There are no invariant curves of T, if and only if there exists a glancing orbit.

PROOF. If there is an invariant curve, there is trivially no glancing orbits because the regions on both sides of the curve are left invariant. Assume now there is no glancing orbit. This means there is an $\epsilon > 0$ such that for all $y_0 < 1 - \epsilon$ we have $y_n > -1 + \epsilon$. Consider the region $Y = \{y < 1 - \epsilon\}$. The set $\bigcup_n T^n(Y)$ is a T-invariant set which does not intersect $\{y > 1 - \epsilon\}$. The boundary of this curve is an invariant curve. (One actually knows that such a curve must be the graph of a Lipshitz continuous function).

THE JACOBEAN. Let κ_i denote the curvature at the impact point and angle θ_i the impact angle and let l_i the length of the path from the impact point x_{i-1} to the impact point x_i. The following formula is well known in geometrical optics and used everywhere in the billiard literature like in the book of Kozlov-Treshchev.

> LEMMA: There are coordinates for which the Jacobean $DT(x_i, y_i)$ of the billiard map has the form
>
> $$B_i = \begin{bmatrix} 1 & 0 \\ -\frac{2\kappa_i}{\sin(\theta_i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & l_i \\ 0 & 1 \end{bmatrix}$$

Remark: This is the composition of the Jacobean belonging to the translation and the Jacobean belonging to the reflection at the wall. The value $g_i = \frac{\sin(\phi_i)}{2\kappa_i}$ is the length of the billiard ball in the circle on the normal to the reflection point which is tangent to the table and has radius $1/(2\kappa_i)$.

PROOF OF THE JACOBEAN FORMULA. The formula can be derived geometrically. Instead, we find an algebraic derivation from the Euler equations. It is still a bit messy.

We use the notation h_1, h_{11} for the first and second partial derivative with respect to the first variable and similar h_{12} for the mixed partial derivative. The billiard map $S : \begin{bmatrix} x_i \\ x_{i-1} \end{bmatrix} \to \begin{bmatrix} x_{i+1} \\ x_i \end{bmatrix}$ is equivalent to the second order recursion $h_1(x_i, x_{i+1} + h_2(x_{i-1}, x_i) = 0$. Differentiation of these Euler equation with respect to x_i, x_{i-1} gives $\partial x_{i+1}/\partial x_i = -b_i/a_i, \partial x_{i+1}/\partial x_{i-1} = -a_{i-1}/a_i$, where

$$a_i = h_{12}(x_i, x_{i+1})$$

and

$$b_i = h_{11}(x_i, x_{i+1}) + h_{22}(x_{i-1}, x_i) \; .$$

The Jacobean of S is

$$dS = \begin{bmatrix} -b_i/a_i & -a_{i-1}/a_i \\ 1 & 0 \end{bmatrix} \; .$$

With a first coordinate transformation $F_i = \begin{bmatrix} a_i^{-1} & 0 \\ 0 & 1 \end{bmatrix}$ we can achieve that the determinant is 1:

$$F_i^{-1} dS F_{i-1} = A_i = (a_{i-1})^{-1} \begin{bmatrix} -b_i & -a_{i-1}^2 \\ 1 & 0 \end{bmatrix} \; .$$

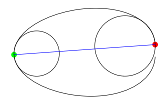
Geometrically, we have

$$a_i = \frac{\sin(\theta_i)\sin(\theta_{i+1})}{l_i}, \; b_i = \sin^2(\theta_i)(\frac{1}{l_i} + \frac{1}{l_{i-1}}) - 2\sin(\theta_i)\kappa_i \; ,$$

where $l_i = h(x_i, x_{i+1})$ are the lengths of the secants, $\theta_i = \theta(x_i, x_{i+1})$ and $\kappa_i = \kappa(x_i)$ are the curvatures at the reflection points. Plugging this in the Jacobean gives with $G_i = \begin{bmatrix} 0 & -\sin(\theta_i) \\ 1/\sin(\theta_i) & \sin(\theta)/l_i \end{bmatrix}$ the new Jacobean

$$G_i^{-1} \cdot A_i \cdot G_{i-1} = \begin{bmatrix} 1 & l_i \\ -\frac{2\kappa_i}{\sin(\theta_i)} & 1 - \frac{2\kappa_i}{\sin(\theta_i)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{2\kappa_i}{\sin(\theta_i)} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & l_i \\ 0 & 1 \end{bmatrix} \; .$$

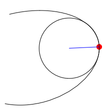
STABILITY OF PERIOD 2 ORBITS. Having the Jacobean given in geometric terms allows to see, whether periodic orbits are stable or not. Inspection of the trace of $B_2 B_1$ (a matrix which is similar to the Jacobean of T^2 and so has the same trace) shows:



LEMMA. Assume ρ_i are the radii of curvature at the impact points. Assume $\rho_1 < \rho_2$. If $l > \rho_1 + \rho_2$ or $\rho_1 < l < \rho_2$, then the periodic orbit of period 2 is hyperbolic. If $l > \rho_2$ or $l < \rho_1 + \rho_2$, it is elliptic.

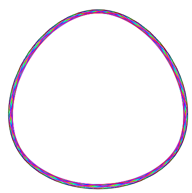
The fastest verification of th lemma is to run a line of Mathematica which gives the trace of the product of the four matrices. For example, the long axis of a non-circular ellipsoid is a hyperbolic periodic point. The short axis is an elliptic periodic point.

CURVATURE. If $r(s)$ is a a curve in the plane parametrized by arc-length, then the curvature $\kappa(t)$ is $|r''(s)|$. If $r(t)$ is the curve given by an arbitrary parameterization, define the unit tangent vector $\vec{T}(t) = \vec{r}\,'(t)/|\vec{r}\,'(t)|$. We get the curvature $\kappa(t) = |\vec{T}'(t)|/|\vec{r}\,'(t)|$. The function $\rho(t) = 1/\kappa(t)$ is called the **radius of curvature**. With the crossed product $(a, b) \times (c, d) = ad - bc$ in two dimensions, we have a more convenient formula $\kappa(t) = \frac{|\vec{r}\,'(t) \times \vec{r}\,''(t)|}{|\vec{r}\,'(t)|^3}$.

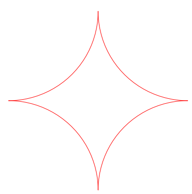


ROLE OF CURVATURE. The curvature of the table plays an important role for the billiard dynamics. Here are some known results:

- Mather has shown that if the table has a flat point, this is a point at which the curvature vanishes like at 4 points of $x^4 + y^4 = 1$, then the billiard map T has no invariant curve at all.
- Lazutkin and Douady have proven using KAM theory that for a smooth billiard table with positive curvature everywhere, there always are "whisper galleries" near the table boundary.



- From Andrea Hubacher (who had obtained this result as an undergraduate student at ETH) is the result that a discontinuity in the curvature of the table does not allow caustics near the boundary. For example, tables obtained by the string construction at a triangle (see homework) do not allow invariant curves near the boundary.
- It is easy to see that billiards for which the table has negative curvature everywhere, the Lyapunov exponent is positive. The Matrices B_i have then positive entries as we will just see.



POSITIVE MATRICES. If we multiply positive matrices with each other, the norm of the product grows exponentially.

LEMMA. If $\det(A(x)) = 1$ for all x and $[A]_{ij}(x) \geq \epsilon > 0$, then the Lyapunov exponent $\lambda(x) = \lim_{n\to\infty} \log ||A(T^{n-1}x)A(T^{k-2}x) \cdots A(x)||$ satisfies $\lambda(A) \geq \frac{1}{2}\log(1 + 2\epsilon^2)$.

PROOF (Wojtkowski). Define the function F on pairs of vectors by $v = (v_1, v_2) \mapsto F(v) = (v_1 \cdot v_2)^{1/2}$. For a matrix B with determinant 1 satisfying $[B]_{ij}(x) \geq \epsilon$, define $\rho(B) = \inf_{F(v)=1} F(Bv)$.

(i) Given a 2×2-matrix A satisfying $[A]_{ij} \geq \epsilon$. Then $\rho(A) \geq (1 + 2\epsilon^2)^{1/2}$. Proof: If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $w = (w_1, w_2)$ with $F(w) = (w_1 w_2)^{1/2} = 1$, then $F(Aw) = (aw_1 + bw_2)^{1/2}(cw_1 + dw_2)^{1/2} \geq (ad - bc + 2bc)^{1/2} \geq (1 + 2\epsilon^2)^{1/2}$.

(ii) $||B|| \geq \rho(B)$. Proof: Take $v = (1, 1)$. Then $||A|| \geq \frac{|Av|}{|v|} \geq \frac{F(Av)}{F(v)} \geq \rho(A)$.

(iii) $\rho(AB) = \inf_{F(v)=1} F(ABv) \geq \inf_{F(Bv)=1} \frac{F(ABv)}{F(Bv)} \cdot \inf_{F(v)=1} F(Bv) = \rho(A) \cdot \rho(B)$.

(iv) We get from (ii),(iii),(i) that $\frac{1}{n}\log ||A^n(x)|| \geq \frac{1}{n}\log(\rho(A(T^{n-1}x) \dots \rho(A(x))) \geq \frac{1}{n}\log((1 + 2\epsilon^2)^{n/2})$.

CLASSES OF CHAOTIC BILLIARDS. Remember that $g = \frac{\sin(\theta)}{2\kappa}$, and l is the length of the trajectory.

THEOREM (Wojtkowski) Assume, a piecewise smooth convex table has the property that for any pair of points x, x', on the non-flat parts of the curve $2g + 2g' \leq l(x, x')$, with strict inequality on a set of positive measure, then the billiard map T has positive Lyapunov exponents on a set of positive measure.

PROOF. The Jacobian matrix is conjugated to $B_2(x)B_1(x)$. A vector $v = (1, f)$ is mapped by the matrix $B_1(x)$ to the vector $(1, f + l(x))$. This vector is then mapped by $B_2(x)$ to the vector

$$(1 - (f + l(x))/2g(Tx), f + l(x))$$

which is after a rescaling of length equal to the vector

$$(1, \frac{(f + l(x))g(Tx)}{2g(Tx) - f - l(x)} .$$

If we don't care about the length of the vector, the map $v \mapsto B(x)v$ is determined by the map

$$K : f \mapsto f + l \mapsto \frac{1}{1/(f + l) - 1/g(T)} = \frac{(f + l)2g(T)}{2g(T) - f - l} .$$

At each point $x \in X$, we define a basis given by $e_2(x) = (1, 0)$ and $e_1(x) = (1, -g(x))$.
Claim: Assume $2g(x) + 2g(Tx) \leq l(x)$ with inequality on a set of positive measure. In this basis, the matrix $B(x)$ is positive and there exists a set of positive measure, where $B(x)_{ij} \geq \epsilon > 0$ for some $\epsilon > 0$ so that we can apply the previous lemma on positive matrices.

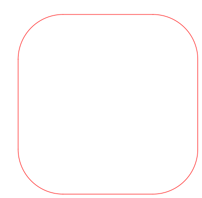
Proof. We have to show that the map K maps the interval $[0, -2g(x)]$ into the interval $[0, -2g(Tx)]$ and into its interior for a set of positive measure because:

$$K(0) = \frac{l(x)2g(Tx)}{2g(Tx) - l(x)} \geq -2g(Tx) .$$

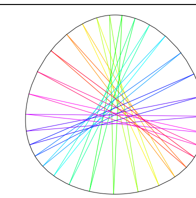
$$K(-2g(x)) = \frac{(-2g(x) + l(x))2g(Tx)}{2g(Tx) + 2g(x) - l(x)} \leq 0 .$$

BUNIMOVICH STADIUM. A famous example is the stadium, where two half circles are joined by straight lines. An other example is the rounded square.

For these billiards, one knows actually much more. They are ergodic and chaotic in the sense of Devaney, a notion we have met earlier in this course. The prove of ergodicity is not so easy. One has to analyze some stable and unstable manifolds and verify that they are dense.

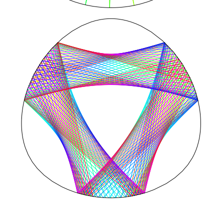


OPEN PROBLEMS. The following problems are open mathematical problems. The first two problems probably go back to Poincaré. The third problem is an old problem in **smooth ergodic theory**. The difficulty of that problem is that for a smooth convex billiards, there are lots of invariant curves and also lots of elliptic periodic orbits consequently, the chaotic regions are mingled well with the stable regions and the techniques described in this handout do not work.



1) Are periodic orbits dense in the annulus for a general smooth Birkhoff billiard?

2) Is the total measure ("area") of the periodic orbits always zero in the annulus? One knows it for period 3 (Rychlik).



3) Does there exist a smooth convex billiards with positive Lyapunov exponents on a set of positive measure (="area"="probability")?